OXFORD

TEACHING EPIDEMIOLOGY

a guide for teachers in epidemiology, public health + clinical medicine

FOURTH EDITION

Edited by **JØRN OLSEN**
NAOMI GREENE
RODOLFO SARACCI
DIMITRIOS TRICHOPOULOS

Teaching Epidemiology

Teaching Epidemiology

A guide for teachers in epidemiology, public health, and clinical medicine

FOURTH EDITION

Edited by

Jørn Olsen

Naomi Greene

Rodolfo Saracci

Dimitrios Trichopoulos

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First edition published (edited by Jørn Olsen and Dimitrios Trichopoulos) 1992
Reprinted 1993
Second edition published (edited by Jørn Olsen, Rodolfo Saracci, and Dimitrios Trichopoulos) 2001
Third edition published (edited by Jørn Olsen, Rodolfo Saracci, and Dimitrios Trichopoulos) 2010
Fourth edition published (edited by Jørn Olsen, Naomi Greene, Rodolfo Saracci, and Dimitrios Trichopoulos) 2015

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2014953045

ISBN 978–0–19–968500–4

Printed in Great Britain by
Clays Ltd, St Ives plc

Preface

Teaching epidemiology requires teaching skills and knowledge. The overriding requirement is knowledge of the subject matter. The general advice is simple: if you are not an expert on a topic, you should enrich your background knowledge before you start teaching. In this book we help you to locate the most important sources of information you need to study before you start.

In addition, we provide expert teachers' advice on how best to structure teaching—what has worked in their hands. You should not, however, expect that these guidelines will automatically work for you. You have to find your own personal style and use examples of relevance for your audience. It is, nevertheless, always useful to make sure your teaching follows a predefined logical sequence. The book will help you to set up this structure.

Most experienced epidemiologists are able to write and present scientific findings by complying with the established rules for scientific writing. Teaching is different because you also have to establish personal contact. Without personal contact, teaching may well be replaced by reading or web-based courses. Personal contact requires that the teacher wants to teach and that the students are willing to learn, or at least to give it a try. Evaluation of your success as a teacher includes assessment of your knowledge and experience as well as how the teaching was received. Any serious evaluation takes both aspects into consideration.

Epidemiology is an old discipline but its concepts and principles are evolving rapidly and, for that reason, a book like this needs frequent updating. Teachers have different ideas as to what the level of sophistication in methods should be and where to focus the attention of the students. We have not sought consensus when inviting the authors to contribute. Science is not driven by consensus but rather by diversity.

We advise you to read and make your own judgements and develop your preferred trajectory. But first, see what older and more experienced colleagues have to offer. Then—but only then—you can throw it away.

This book is a fourth edition of *Teaching Epidemiology*, first published in 1992. The content has changed substantially since the second version (2007). The first edition was published by Oxford University Press and the Commission of the European Communities. Later versions are published by Oxford University Press in collaboration with the International Epidemiological Association

and the European Educational Programme in Epidemiology. The second edition of the book was awarded 'Highly Commended in the Basis of Medicine Category' at the 2002 BMA Medical Book Awards in London.

We would like to thank all authors who agreed to share their experience and knowledge with their less experienced colleagues. Without their contributions there would not have been any book. We also thank Pernille Kümpel and Rikke Sinding for their important administrative support and technical skills.

Jørn Olsen
Naomi Greene
Rodolfo Saracci
Dimitrios Trichopoulos

Acknowledgements

Dimitrios Trichopoulos died 1 December 2014 at 75 years of age. We thank him for his valuable contribution to this book over the years and for his generosity as a teacher, scientist, and friend.

Contents

Contributors

J. H. Abramson
Professor Emeritus of Social Medicine, School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Israel

Anders Ahlbom
Professor of Epidemiology, Chair and Director, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Per Kragh Andersen
Professor, Institute of Public Health, Dept. Of Biostatistics, University of Copenhagen, Denmark

Kelly K. Anderson
Assistant Professor, Department of Epidemiology and Biostatistics, University of Western Ontario, Canada

Josep M. Antó
Director and Researcher, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain

Constantine N. Antonopoulos
Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Greece

Vibeke Baelum
Professor, Department of Dentistry, Health, Aarhus University, Denmark

John A. Baron
Professor of Medicine, University of North Carolina at Chapel Hill, USA

Olga Basso
Associate Professor, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University Faculty of Medicine, Montreal, Quebec, Canada

Yoav Ben-Shlomo
Professor, School of Social and Community Medicine, Bristol, UK

Tarun Bhatnagar
Scientist C, National Institute of Epidemiology (Indian Council of Medical Research), Tamil Nadu Housing Board, Ayapakkam, Chennai, India

Jakob Bue Bjørner
Professor, National Research Centre for the Working Environment, Copenhagen, Denmark

Harry Campbell
Professor, Centre for Population Health Sciences and Institute of Genetics and Molecular Medicine, College of Medicine and Vet Medicine, University of Edinburgh, UK

Ulisses Confalonieri
Professor, Laboratory on Education for Health and the Environment (LAESA), Rene Rachou Research Center, The Oswaldo Cruz Foundation (FIOCRUZ), Brazilian Ministry of Health, Belo Horizonte, Brazil

Paul Elliott
Professor Small Area Health Statistics Unit, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, UK

Betsy Foxman
Professor, Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA

Rebecca Fuhrer
Professor, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University Faculty of Medicine, Montreal, Quebec, Canada

Patricia Gerakopoulou
Research Associate, National and Kapodistrian University of Athens Medical School, Department of Hygiene, Epidemiology and Medical Statistics, Athens, Greece

A. L. Golbeck
Professor, The School of Public and Community Health Sciences, The University of Montana, Missoula, MT, USA

Naomi Greene
University of California, Los Angeles, Fielding School of Public Health, USA

Anna Hansell
Assistant Director, Small Area Health Statistics Unit, MRC-PHE, Centre for Environment and Health School of Public Health, Imperial College London, UK

Susan Jick
Professor of Epidemiology, Director, Boston Collaborative Drug Surveillance Program, Boston University School of Public Health, MA, USA

Nancy Krieger
Professor of Social Epidemiology, Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA, USA

Diana Kuh
Professor, MRC Unit for Lifelong Health and Ageing, London, UK

Pagona Lagiou
Professor of Hygiene and Epidemiology, University of Athens Medical School and Adjunct Professor of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Marc Lipsitch
Professor of Epidemiology, Department of Epidemiology and Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

Olli S. Miettinen
Department of Epidemiology, Biostatistics, and Occupational Health, McGill University Faculty of Medicine, Montreal, Quebec, Canada

C. A. Molgaard
Professor and Chair, The School of Public and Community Health Sciences, The University of Montana, Missoula, MT, USA

Ellen Aagaard Nøhr
Professor, Department of Obstetrics and Gynaecology, Odense University Hospital and University of Southern Denmark, Odense, Denmark

Jørn Olsen
Professor, Section for Epidemiology, Department of Public Health, Aarhus University, Denmark and Adjunct Professor of Epidemiology, University of California, Los Angeles, USA

Neil Pearce
Professor, London School of Hygiene and Tropical Medicine, UK

Eleni Petridou
Professor of Preventive Medicine and Epidemiology, National and Kapodistrian University of Athens, Medical School, Department of Hygiene, Epidemiology and Medical Statistics, Greece

Kenneth J. Rothman
Distinguished Fellow, RTI Health Solutions, Research Triangle Institute, Research Triangle Park, NC, USA

John F. Rothrock
Director, Renown Neurology Clinic, Professor and Chief, Division of Neurosciences University of Nevada and Reno School of Medicine, USA

Rodolfo Saracci
Adjunct Professor, Section for Epidemiology, Department of Public Health, Aarhus University, Denmark

Susan Service
UCLA Semel Institute for Neuroscience and Behavior, University of California, Los Angeles, USA

Shilu Tong
School of Public Health, Institute of Health and Biomedical, Innovation Queensland University of Technology Kelvin Grove, Qld. 4059, Australia

Harold C. Sox Jr
Professor Emeritus, The Geisel School of Medicine at Dartmouth, Hanover, NH, USA

Sherri O. Stuver
Assistant Director, Department of Epidemiology, Boston University School of Public Health, MA, USA

Henrik Toft Sørensen
Professor, Department of Clinical Epidemiology, Aarhus University Hospital, Denmark

Dimitrios Trichopoulos
Vincent L. Gregory
Professor of Cancer Prevention and Professor of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Georgios Tsakos
Department of Epidemiology and Public Health, University College London, UK

Walter C. Willett
Professor, Departments of Nutrition and Epidemiology, Harvard School of Public Health, and Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Part 1

Context

Chapter 1

Introducing the history of epidemiology

Rodolfo Saracci

Nature cannot know its own history; humans can.

Introduction to the history of epidemiology

Why teach the history of epidemiology?

'Know yourself': this Socratic maxim expresses the rationale for learning the history of epidemiology. Self-knowledge, as derived from a view of the development in time of epidemiology, promotes a positively critical attitude towards the discipline and its practice by fostering an appreciation of

- the common features of all branches of epidemiology as a population health science beyond today's subdivision in specialized areas;
- the relationship of epidemiology to other scientific disciplines, and its methodological specificities, strengths, and weaknesses in respect to them;
- the process of the emergence of key concepts, be they methodological, such as 'risk', or substantive, such as the modes of diffusion of pathogens in the community: this cumulative but irregular accrual of knowledge is characterized by controversies, blind alleys, sheer errors and, not rarely, also by material hurdles, as well as personal and institutional conflicts; it usually appears long when time is measured in years but much less so when a 'generation' of scientists is—more appropriately—taken as the time unit;
- the influence of the demographic, health, social, cultural, and economic context on the development of epidemiology and epidemiological methods;
- the role of epidemiology in society through its impact in the health field, an impact largely mediated through the essential functions of epidemiology within public health; and

- the roots and dynamics of present trends in epidemiology, and the options for reinforcing, inflecting, or contrasting them; in particular, the pressure for resources and research to be concentrated on the 'theme of the day', ignoring or paying lip service to the past and glossing over future implications. The increasingly technical and specialized character of most epidemiology textbooks does little to keep this trend in check.

Each of these six perspectives on epidemiology should be regarded as essentially descriptive and only tentatively as explanatory. Presentations on the long-term evolution of epidemiology, especially 'bird's eye views', as the one in 'Annex: an historical sketch', are affected to a variable but usually substantial extent by 'teleological bias' (Saracci 2011). The bias arises from the unavoidable tendency to reconstruct sequences of and links between past ideas, concepts, and events in such a way that they can logically and consistently account for today's situation, for instance, today's concept of 'risk factor'. This in fact generates not a causal but a finalistic (pseudo) explanation because, as the historian Trevor-Roper noted (Pearl et al. 1981), history is not only what occurred but what occurred in the context of what could have occurred, namely, in light of all possible alternative paths of evolution. I am not aware of any systematic work that employs this counterfactual approach in the study of the long-term development of epidemiology; indeed, scholarly historical work avoids overall narratives (counterfactual or not) and mostly develops meticulous 'worm's eye views' on limited, specific issues, based on the study of primary documentary sources.

Whom to teach?

The teaching as here outlined is primarily addressed to students pursuing master's or doctoral degrees and who intend to become full-time epidemiologists, as well as to those who will make a large use of epidemiology in their professional work (e.g. public health practitioners, clinicians, and occupational physicians).

This teaching can only be regarded as preliminary for students who may wish to proceed in one of two directions: historical epidemiology as a description and analysis of health and diseases in given areas and past periods of time; and history of epidemiology as the reconstruction of the development of theories, concepts, methods, and practices, including the study of the role of individuals and institutions.

For undergraduate students, for example in medicine, to whom a twenty- to forty-hour course in epidemiology or epidemiological methods is imparted at a number of medical schools, a short historical overview (one or two lectures)

should be offered. It may focus on such critical passages in the rise of epidemiology as the confluence of medical, demographic, and quantitative thinking and the genesis of the Koch's postulates (see 'Annex: an historical sketch') as well as deal with the societal aspects of the modern history of tobacco and health.

How and when to teach?

The chapter is a guide to the preparation of a short introductory module of eight to ten hours within a master's or doctoral programme; it does not provide a script for lectures. In many epidemiology programmes, historical notes are confined to occasional (if any) comments within the body of other modules: this module aims instead at presenting structured material as an individualized teaching block to students, provided the students have already studied at least one or, better, more modules in general epidemiology and epidemiological methods. An important point to which the teacher should pay attention are features in the development of epidemiology particular to his/her country: an effort should be made to take these into account, lest elements essential to the understanding of the present situation are missed.

Teaching objectives

The objective of the module is to focus and raise the motivation of students for the historical perspective in epidemiology, namely for the exercise of asking and answering, when confronted with a methodological or substantive issue such questions as: how did the issue develop overall? Was it recognizable and what form was it in at, for instance, the beginning of the twentieth century? What were its antecedent and related issues? How was it tackled within the frame of existing knowledge? Did it induce critical advancements in knowledge? Was it instrumental for the development of approaches and methods that are more widely applicable? Did any social, economic, political, or ethical factors have a non-trivial influence on it? Were any social, economic, political, or ethical consequences derived from its treatment within epidemiology? What can an historical exploration tell us about the present status and future evolution of the issue?

Similar questions arise in daily work, typically when reviewing the literature on a specific research topic. Looking at them historically means expanding questions and answers in three ways:

- in time, going back not just a few years or a couple of decades (although this may be perfectly adequate for the strict needs of the research topic at hand);
- in extent, tracing the connections of the issue internally within epidemiology and other sciences and externally to society; and

- in viewpoint, as these connections can set the issue in the context of various viewpoints rather than making it an element within a unique pattern of causal relationships. In fact, if the very concept of cause is problematic in physics and if the causal nature of associations turns out often to be in doubt in epidemiology, in history it is impossible to identify complex patterns of causes in a unique and definitive way. One reason is that historical knowledge changes in time, not only because of cumulative advances—as in all theoretical or empirical sciences—but also because the point of observation of the historical past moves continuously forward with time, and the landscape that becomes visible to the eye of the observer changes in consequence: history, as a reconstruction of the past, is to a non-negligible extent a function of the present and of what the present allows or does not allow the observer to discern. This unavoidable bias may be grossly amplified when the past is reconstructed not for the purpose of knowing it or trying to explain the present but for justifying particular aspects and views of the present: for instance, rudimentary and lopsided reconstructions of mankind's evolutionary history have been abundantly, and tragically, employed to support racist theories and practices.

Structure of the course

The module includes five sessions:

1. Overview of the history of epidemiology: one lecture with a brief discussion in plenary (one or two hours).
2. The modern history of tobacco and health—scientific aspects: one lecture with an extended discussion in plenary (two hours).
3. The modern history of tobacco and health—societal aspects: one lecture with an extended discussion in plenary (two hours).
4. A paper from the past: reading followed by an extended discussion in plenary (two hours).
5. Present trends linking the past and future: reading of papers followed by an extended discussion in plenary (two hours).

The total time required for the five sessions is nine or ten hours. Papers for reading (in sessions 4 and 5) can be distributed in advance. A class of more than ten to fifteen students should be split into smaller subgroups: the papers' reading can be completed, with some discussion, within subgroups in preparation for a plenary session in which each subgroup presents for general discussion issues identified as important. A written outline of each part with essential references should be prepared for distribution.

Teaching contents

Material for the five components of the module can be found and selected for assemblage and teaching in

- 'Annex: an historical sketch', which also provides a frame for the whole module;
- the references listed at the end of the chapter; these references should be in any case consulted by the teacher, being a primary source of material for all parts of the module; and
- the additional references reported at the chapter end.

Overview of the history of epidemiology

The section entitled 'Annex: an historical sketch' can be used as a basis for this overview. A valuable, fine analysis of the emergence of epidemiological concepts and methods, although not exempt from teleological bias, is contained in Part 1 of *A History of Epidemiologic Methods and Concepts* (Morabia 2004). The interest of the students is heightened if examples are produced from the national and local context. The bulk of significant scientific advances in epidemiology has been concentrated in a relatively small number of countries. However, isolated but important discoveries—often related to special traits of local health and diseases—have come from many more countries and there is virtually no country in which the echo of scientific advances has not been received in some form. It is these aspects which can be exploited for illustrative purposes.

The modern history of tobacco and health: scientific aspects

The identification of the causal role of tobacco smoking in a variety of diseases is a prime success of the new, post-Second World War epidemiology. The historical sketch (see 'The new epidemiology') contains an outline that should be fleshed out using material from the quoted sources. The focus can be on the controversy concerning the etiological role of tobacco smoking, in particular in lung cancer, and on the emergence, under the stimulus of the controversy, of methodologies for data analysis and of criteria for inferring causation in epidemiology. These aspects are covered in *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service* (US Department of Health, Education, and Welfare 1964), especially in chapters 8 and 9, as well as in some of the publications cited in that report and which were written by eminent critics of the smoking causal hypothesis, such as R. A. Fisher, a founder of modern statistical methodology, and J. Berkson, a leading medical statistician.

The modern history of tobacco and health: societal aspects

The direct link between epidemiology and prevention is commonly stated as one of the merits of the epidemiological approach: once causation is established, as it was in the early sixties between tobacco smoking and various pathological conditions, preventive actions can be implemented even without knowledge of the biological mechanisms leading to the condition. This logical sequence often breaks in practice, as the transition from sufficient knowledge for action to actual action is strongly influenced by a number of social, economic, cultural, and psychological factors. First, simply comparing in time the local or national tobacco-smoking patterns and trends with the available established knowledge on health effects and with legislative and health-promoting actions is instructive. Second, as documented and discussed by Proctor (1999), the crucial role of factors other than sound science and epidemiology is highlighted by the case of Nazi Germany, where vigorous anti-smoking campaigns were conducted, inspired by strong ideological reasons. Moreover, informative epidemiological evidence gathered during the Nazi period went ignored till recently for reasons mostly foreign to science. Third, the realization that adverse health effects, particularly lung cancer, result from passive exposure to environmental tobacco smoke (ETS; Hackshaw et al. 1997; Boffetta et al. 1998) has radically changed the prospects of preventive measures; unlike active smoking, which is (partially) voluntary, ETS exposure is wholly involuntary and there is general agreement that involuntary exposure should be controlled mainly through legal restrictions. This prompted researchers to rush in and attack the tobacco–health relationship from this new angle, while at the same time tobacco producers embarked on a major campaign denying that there were any material health effect from ETS. This campaign was often fought using organized and unscrupulous means, as documented by Ong and Glantz (2000). This example offers a lesson on how powerful and well-structured economic interests can lead to the misinterpretation of scientific evidence to the point of active disinformation, and on how inherently liable to distortion by extrascientific forces may be the judgement of experts who think of themselves as independent (Maggi 2000). An additional perverse effect of this polluted climate of debate is that legitimate and illegitimate scientific doubts may become hardly distinguishable.

A paper from the past

A paper addressing a substantive epidemiological issue with the methodological armamentarium available at the time of writing will provide a measure of the methodological gap separating the past from the present. Papers can be selected from the bibliography; or, as an example, the article (cited among 'Additional references') on sickness and stress in operational flying in the Royal

Air Force during the Second World War (Reid 1948) can be chosen. In reading it, students should focus in particular on (a) key issues of design, such as whether and how issues of confounding, bias, and chance were dealt with at the design and analysis stages and how inferences about causation were developed; and (b) how they would design today, within the limits of the data available to the author, a study addressing the same issue of sickness and stress in operational flying and how they would conduct and analyse such a study. From a comparison of (a) and (b), the conclusions in this article (or in any other paper selected for the exercise) may be assessed as not credible, either at the time of writing or today; credible, to a degree that one may even wish to specify (e.g. as moderately credible, credible, or highly credible), at the time of writing but not today, in light of new methodological criteria judged as non-dispensable; or credible at the time of writing as well as today, notwithstanding the inherent limitations deriving from the past status of the art.

Present trends linking past and future

Two sources of readings can be used for this reading and discussion session (the topic of which is discussed in the sections 'Today: epidemiology in the making' and 'Tomorrow's horizon' in 'Annex: an historical sketch'): (1) the series of papers published on the occasion of the turn of the century in a supplement of *International Journal of Epidemiology* (Adami and Trichopoulos 1999; Holland 1999; Hunter 1999; Pearce 1999; Saracci 1999; [Saracci] 1999a, b; Susser 1999; Wall 1999), as they provide a spectrum of views on the evolution of epidemiology; and (2) the chapters in *The Development of Modern Epidemiology: Personal Reports from Those Who Were There* (Holland et al. 2007), as they cover both general and national historical themes. Two or three of these can be distributed to students; each student should read them and then extract from each of them two lists of key points: one containing points with which the student agrees, and another containing points with which the student disagrees. These points, together with supporting arguments, should then be discussed in plenary.

An alternate method is to ask each student (before distributing papers) to prepare an outline of how he/she sees the likely and desirable evolution of epidemiology as a scientific endeavour and as a practice within society. The outlines can then be presented to the class for discussion and compared with published essays.

Areas for consideration when looking from past and present into the future come under three main headings:

- information sources (items such as population disease registries and stores of biological specimens: organization, access, use, protection, ethical and legal implications);

- methodology (items such as methods to improve exposure and endpoints assessment; new study designs; multilevel analyses; pooled data analyses and meta-analyses; risk modelling); and
- aims and uses of epidemiology (items such as the role of epidemiology in clinical medicine, environmental and occupational health, public health, prevention, and the reduction of inequalities in health between and within countries; priorities for research; the communication of results and the effect of indecision on the part of epidemiologists; and the responsibilities of epidemiologists).

Assessing students' achievements

As stated, the object of the module is to focus and raise the interest and motivation of students to gain an historical perspective on epidemiological themes, not to develop the skills necessary to gain this perspective (this would require a more sustained didactic endeavour). Students should be asked via a mini-questionnaire how they rate (on a 'poor, fair, good, excellent' scale) the module in respects of four aspects: increase in knowledge; usefulness for work; stimulation of interest for the historical perspective; and motivation for personally engaging in historical exercises.

Annex: an historical sketch

To review the 'yesterday, today, and tomorrow' of epidemiology (here seen with a bias towards Europe), it may be convenient, though somewhat arbitrary, to consider three periods: early epidemiology, extending from the fifth century BC to around 1830; classical epidemiology, from around 1830 to the 1940s; and new epidemiology, from the 1940s to the present. The full set of quoted references is available in the third edition of this book (Saracci 2010).

Yesterday: a bird's eye view

Early epidemiology

This long stretch of time ran for more than two millennia, from Hippocrates (*c.*470–*c.*400 BC) to the early third of the nineteenth century. Numerous and keen epidemiological observations were made and have been handed down in surviving documents; such observations were based at best on simple or crude methods of investigation (if unfairly judged by our contemporary methodological standards). Epidemiological theories were also elaborated to explain the spreading of diseases, notably those recurrently striking and decimating populations ('epidemics').

On the medical line of development, Hippocrates not only provided concise, accurate, and complete descriptions of actual clinical cases (including cases of diseases such as tetanus, typhus, and phthisis) which remain 'without parallel till the late seventeenth century' (Singer and Underwood 1962) but—as a seminal environmental scientist—especially in his book *On Airs, Waters and Places*, clearly identified the general dependence of health, not on magical influences, but on an identifiable array of natural external factors (Hippocrates 1923).

The reawakening of clinical observation in the seventeenth century, epitomized by the 'English Hippocrates', Thomas Sydenham (1624–89), also brought attention back to the circumstances surrounding the occurrence of a clinical case, thus not only reviving the Hippocratic tradition, but adding to it. In 1700, Bernardino Ramazzini (1633–1714) wrote the following in his *De morbis artificum diatriba*:

> Hippocrates states in *De affectionibus*: 'When you face a sick person you should ask him from what he is suffering, for what reason, for how many days, what he eats and what are his bowel movements.' To all these questions one should be added: 'What work does he do?' (Ramazzini 1982)

Besides being an acute clinician, Ramazzini moved from the observation of individual cases to the consideration of similar cases sharing work circumstances. Thus, he is today regarded as the founder of occupational medicine, a key section of the larger field of environmental medicine and epidemiology. Contemporary with Ramazzini was Giovanni Maria Lancisi (1654–1720), an anatomist and clinician, whose *De subitaneis mortibus* (1707; Lebowitz 1970), in which he reports a detailed pathological investigation of a series of cases of sudden deaths in Rome, is probably the first epidemiological study of a non-communicable condition (Lebowitz 1970). The study was commissioned by the Pope, to whom Lancisi was the personal physician (there were no forms of grant applications in those days!).

On a rather different, essentially demographic line, are the developments such as those taking place already in the late Middle Ages and in Renaissance Italy, in the latter part of the fourteenth century and in the fifteenth century, where, for instance, in Florence and Venice, the counting of deaths and some early form of death certification specifying the cause in broad terms (e.g. 'plague' or 'not plague' during such epidemics) were current and established practices (Carmichael 1986). A major step forward from recording, counting, and accounting to a quantitative analysis of the data was the later accomplishment in London of John Graunt (1620–74), who can be regarded as the founder of demography. *Natural and Political Observations upon the Bills of Mortality* (1662; Dupaquier and Dupaquier 1985) was based on a series of weekly bills

covering individual deaths and their causes in the London area back to 1603. His treatment of the data included three key innovations (Dupaquier and Dupaquier 1985): a critical examination of the sources, attempting to address issues of biased recording; the use of relative frequencies, for example, of deaths, and ratios rather than absolute numbers in his analysis, allowing several correct comparisons to be made; and the application of these methods to tackle concrete problems; the resulting data prompted him to conclude, for instance, that homicides were indeed rather rare; that mortality in the first year of life was higher in males than in females, thus compensating for the slightly higher number of males at birth; and that chronic conditions were killing more people than were acute conditions, apart from the plague. After Graunt, demographic studies progressed with the invention of the first empirical life tables (E. Halley, 1656–1742), while, particularly in France, mathematical tools were being developed for dealing with chance events and probabilities initially arising out of games; such tools were soon seen as equally applicable to the study of such collective phenomena as births, deaths, etc.

The third line, theorization, especially about the fact that the most frequent and murderous diseases appeared obviously 'communicable', either from person to person or from fomites, has a forerunner in the Latin poet, Lucretius (first to second century BC); in his poem *De rerum natura* he hints that 'seeds' of disease can pass from a sick to a healthy individual. It was, however, only much later that Gerolamo Fracastoro (1478–1553), in his *De contagione et contagiosis morbis et eorum curatione* (1546; Winslow 1980), presented the first clear and coherent germ theory of disease, 'a mountain peak in the history of etiology, perhaps unequalled by any other writer between Hippocrates and Pasteur (Winslow 1980). Fracastoro theorized that a variety of diseases are caused by transmissible, self-propagating entities (germs) which, however, were conceived as substances more akin to present-day viruses than to bacteria. Correctly, he thought that these agents were specific to each disease and could spread person to person or through infected articles (fomites) or at a distance. He went as far as arguing that treatment should consist either of the destruction of the germs by such processes as heat or cold (which is obviously correct), the evacuation of the germs from the body, halting the putrefaction processes caused by these germs, or neutralizing them by antagonistic substances (which again was correct but, unfortunately, these were not available). He is also on record as having not only described but also given the name to a new disease making ravages in his time—syphilis (1530).

The three streams in early epidemiology—medical, demographic, and theoretical—coalesced in an effective way only towards the end of the eighteenth century and the beginning of the nineteenth century, giving rise to

epidemiology as we recognize it today, an investigation of diseases and their etiology at the population level. What had been missing during the very long early phase was not so much the individual components of the epidemiology approach as the integrated and systematic process of empirical observations, quantitative description, hypothesis formulation, deductive reasoning, and empirical testing on new observational or experimental data which started in science with Galileo Galilei (1564–1642) at the beginning of the seventeenth century and gradually spread from physics to other branches of study. In biology, an early high point in this combination of observation, experiment, and quantitative reasoning was the discovery of the circulation of blood by William Harvey (1578–1657), a contemporary of Galilei.

Classical epidemiology

With the advent of the industrial transformation of western Europe, starting in Great Britain and propagating from the mid eighteenth century to the continental countries in the next decades, 'crowd diseases' emerged which struck the populations amassed in the slums of the fast-growing centres of industrial development: London, Glasgow, Manchester, Paris, Lyon, Berlin, etc. This provided the decisive stimulus and at the same time the observational field for epidemiology, which developed as the investigation facet of a vast public health movement. Only a few landmarks and figures can be briefly cited here.

In Great Britain, medical registration of deaths had been introduced in 1801 and, in 1838, William Farr (1807–83) introduced a national system of recording causes of death. Once the mechanism started to work, it provided a wealth of data which Farr himself first analysed with great skill, making full use of life-table techniques, close in most details to those in present-day use, and procedures for standardizing rates. He was also instrumental in building up a classification of diseases for statistical purposes, both national and international. His analyses, published from the Registrar General's Office at regular intervals, gave a picture of the evolving health condition of the population of Great Britain; this picture subsequently drew the attention of all the social investigators of the Victorian period, including Marx and Engels.

The work of John Snow (1813–58), a contemporary of William Farr, is generally cited as an example of a brilliant analytical investigation which led to both the identification of a pathogenic agent and the elimination of that agent from the environment. Cholera (Asiatic cholera) had started to rage in India and then moved westwards, the first epidemic hitting Great Britain in 1831–2 and causing at least 60,000 deaths. Snow directly investigated the subsequent major epidemic episodes in London in 1849 and 1854, focusing attention on the role that polluted water might have played in the spread of the disease. Among a

Table 1.1 Mortality from cholera in 1849 and 1854 in London areas supplied by the Southwark and Vauxhall Waterworks Company and/or the Lambeth Waterworks Company

Water supply company	Number of deaths attributed to cholera	
	1849	1854
Southwark and Vauxhall	Article I. 2,261	2,458
Both companies	3,905	2,547
Lambeth	162	37

Adapted from John Snow, *On the Mode of Communication of Cholera*, John Churchill, New Burlington Street, London, UK. Copyright © 1855 John Churchill.

number of other observations, he noted (Table 1.1) that while roughly the same number of deaths had occurred in 1849 and 1854 in London districts supplied by the Southwark and Vauxhall Waterworks Company, there were markedly fewer deaths in 1854 than in 1849 in those districts supplied by the Lambeth Waterworks Company. No major change in population had occurred between 1849 and 1854; however, unlike Southwark and Vauxhall, Lambeth had changed its source of water supply during that time by moving higher up the Thames to a location probably above, as Snow conjectured, the greatest sources of contamination by the city sewage.

Indeed, when he computed the death rates from cholera (Table 1.2), they were more than twenty times lower for the districts supplied by Lambeth than for those supplied by Southwark and Vauxhall. Strong corroboration of these findings came from a more refined investigation: in some areas, the water supplies for the two companies happened to be closely intermixed, with some houses receiving their water from Lambeth and others from Southwark and Vauxhall. The number of houses and the size of the pertinent populations belonging to each company were known, but a door-to-door inquiry was needed and was indeed carried out by Snow on all cholera cases in order to ascertain to which company the water supply of their homes belonged. This allowed the correct calculation of valid rates of cholera occurrence. The results are shown in Table 1.3 and clearly demonstrate that, even within the same physical area, the origin of the water supply separates in a clear-cut way populations with high and low rates of disease occurrence. All of these observations can be seen, in fact, as the test of a lucid theory of the etiology of communicable diseases that Snow had elaborated and presented in an 1853 paper titled 'On continuous molecular changes', along the line of previous work by the German pathologist, Jacob Henle (1809–1885; Winkelstein 1995).

Table 1.2 Mortality from cholera in London districts supplied by the Southwark and Vauxhall Waterworks Company and/or the Lambeth Waterworks Company, 8 July to 26 August 1854

Water supply company	Districts and subdistricts	Population in 1851	Deaths from cholera	Deaths from cholera per 1,000 population
Southwark and Vauxhall	St Saviour, Southwark	19,709	125	6.3
	St Olave	8,015	53	6.6
	St John, Horsleydown	11,360	51	4.5
	St James, Bermondsey	18,899	123	6.5
	St Mary Magdalen	13,934	87	6.2
	Leather Market	15,295	81	5.3
	Rotherhithe	17,805	103	5.8
	Wandsworth	10,560	54	5.1
	Battersea	9,611	11	1.1
	Putney	5,280	1	0.2
	Camberwell	17,742	96	5.4
	Peckham	19,444	59	3.0
Total for districts supplied by Southwark and Vauxhall		**167,654**	**844**	**5.0**
Both companies	Christchurch, Southwark	16,022	25	1.6
	Kent Road	18,126	57	3.1
	Borough Road	15,862	71	4.5
	London Road	17,836	29	1.6
	Trinity, Newington	20,922	58	2.8
	St Peter, Walworth	29,861	90	3.0
	St Mary, Newington	14,033	21	1.5
	Waterloo Road (1st)	14,088	10	0.7
	Waterloo Road (2nd)	18,348	36	2.0
	Lambeth Church (1st)	18,409	18	1.0
	Lambeth Church (2nd)	26,748	53	2.0
	Kennington (1st)	24,261	71	2.9
	Kennington (2nd)	18,848	38	2.0
	Brixton	14,610	9	0.6
	Clapham	16,260	24	1.5
	St George, Camberwell	15,849	42	2.7

(*continued*)

Table 1.2 (continued) Mortality from cholera in London districts supplied by the Southwark and Vauxhall Waterworks Company and/or the Lambeth Waterworks Company, 8 July to 26 August 1854

Water supply company	Districts and subdistricts	Population in 1851	Deaths from cholera	Deaths from cholera per 1,000 population
Total for districts supplied by both companies		**300,149**	**652**	**2.2**
Lambeth	Norwood	3,977	8	2.0
	Streatham	9,023	6	0.7
	Dulwich	1,632	0	0.0
	Sydenham	4,501	4	0.9
Total for districts supplied by Lambeth		**19,133**	**18**	**0.9**

Adapted from John Snow, *On the Mode of Communication of Cholera*, John Churchill, New Burlington Street, London, UK. Copyright © 1855 John Churchill.

In France, the influence of the great mathematicians such as D'Alembert, Condorcet, the Swiss Euler and Bernoulli families, Lagrange, and Laplace, all of whom worked to various extents on probability and statistics during the eighteenth and early nineteenth centuries, was strongly felt in the medical field. A central figure in this development was the physician Pierre Louis (1787–1872), who introduced the 'numerical method' in medicine and produced statistical evidence that the then widespread practice of bloodletting was virtually ineffective or even dangerous. That the scientific climate in the first half of the nineteenth century had become favourable to a quantitative study of medical phenomena is shown by the substantial number of articles published in most of the European countries dealing with problems in quantitative biology or in the clinical or public health domains (Buck et al. 1988). Even in a country like Italy, which by that time had become, after an illustrious past, rather peripheral in scientific development, one finds evident traces of this atmosphere. For instance, at the first Congress of Italian Scientists, Pisa 1838 (*Riunione degli scienziati italiani* 1939), it was proposed that, in order to compare different treatments, the best method would be to administer them in different wards of large hospitals to which access of patients would be on a strict rotation basis, without any possibility of choice on the part of the physicians; the outcome of each treatment would then be carefully recorded and counted, and the whole process, as well as the interpretation of the results identifying the superior treatment (if any), would be strictly monitored and reported by a steering committee.

Table 1.3 Mortality from cholera in London districts supplied by the Southwark and Vauxhall Waterworks Company and/or the Lambeth Waterworks Company, 8 July to 26 August 1854, according to company supplying the water for individual houses

Company supplying water to individual houses	Districts and subdistricts	Population in 1851	Deaths from cholera	Cholera death rate per 1,000 population
Southwark and Vauxhall	St Saviour, Southwark	19,709	125	6.3
	St Olave	8,015	53	6.6
	St John, Horsleydown	11,360	51	4.5
	St James, Bermondsey	18,899	123	6.5
	St Mary Magdalen	13,934	87	6.2
	Leather Market	15,295	81	5.3
	Rotherhithe	17,805	103	5.8
	Wandsworth	10,560	54	5.1
	Battersea	9,611	11	1.1
	Putney	5,280	1	0.2
	Camberwell	17,742	96	5.4
	Peckham	19,444	59	3.0
	Southwark and Vauxhall	**167,654***	**738**	**4.4**
Lambeth	Norwood	3,977	8	2.0
	Streatham	9,023	6	0.7
	Dulwich	1,632	0	0.0
	Sydenham	4,501	4	0.9
	Lambeth	**19,133***	**4**	**0.2**
Southwark and Vauxhall	**Districts supplied by both companies**	**98,862**	**419**	**4.2**
Lambeth	**Districts supplied by both companies**	**154,615**	**80**	**0.5**
	Rest of London	**1,921,972**	**1,422**	**0.7**

Adapted from John Snow, *On the Mode of Communication of Cholera*, John Churchill, New Burlington Street, London, UK. Copyright © 1855 John Churchill.

* Overestimated by a small amount, since this figure includes population with no water supply.

The highest degree of synthesis between experimental science, medicine, simple but penetrating demographic investigation, and public health concern was probably achieved in the unique personality of the German scientist, Rudolf Virchow (1821–1902). His work in pathology is regarded as a cornerstone of modern medicine: not only is he the acknowledged founder of (microscopic) cellular pathology, but he also wrote and was very active in the field of public health, inspired by his belief that 'medicine is a social science'. It is interesting to see how the flow of communication was taking place at that time and how, for instance, an agreed system of classification of diseases, basic to any epidemiological work, was being shaped up. At the International Congress of Statistics in 1985, Rudolf Virchow (1985) stated

> The form of the bulletin indicated by Mr. Farr can be recommended from the practical and medical point of view, because it contains one column for the disease, and another for the consequences of the diseases that have been the immediate cause of death.

One can clearly recognize here the basic concept and structure of current death certification, separating underlying causes from proximate causes, as well as the separate classification of accidents and traumatic events even now present in the International Classification of Diseases (ICD). It fell to another German scientist, Robert Koch (1843–1910), in the wake of the fundamental discovery of microorganisms by Louis Pasteur (1822–95), not only to contribute to the discovery of the agents of several diseases (including the actual identification of the major agent for tuberculosis), but also to formulate a set of criteria for establishing causality in epidemiological studies. With the new ability to isolate from healthy and diseased people a wide variety of microorganisms, the entirely new problem arose at that time of sorting out the few capable of causing a disease from the majority of innocent passengers. Koch's (1893) criteria, among others, addressed this issue in a sharp way, stating that, in order to be regarded as a causative agent of a disease, a microorganism:

1. should be found in all subjects with the disease;
2. should not be found as a fortuitous agent in other diseases; and
3. should be recovered from the body, grown in pure culture, and be capable of reproducing the disease in some animal species.

While the first criterion is formulated in a strictly deterministic way and therefore looks at a glance radically different from our contemporary probabilistic concepts, one may doubt whether it has ever been applied as such without in practice making allowance for a margin of error in what one would regard as 'all subjects' (99%? 95%?). It is the third criterion which more sharply differs from those nowadays quoted in the literature and which follow the guidelines put

forward by A. B. Hill (1965), as these include as one element supporting causality 'the biological plausibility'. This is much weaker and less strict than the ability to reproduce the disease in some (i.e. one or more) uninfected and susceptible animal species, which reflects an attitude of giving full weight to the result of experiments in animals, a feature which became somewhat blurred with the advent of the 'new epidemiology' in the 1940s. Indeed, while the further evolution of epidemiology after Pasteur and Koch throughout the last part of the nineteenth century and the first part of the twentieth century largely occurred hand in hand with parallel experimental and laboratory developments in the field of microbiology, the new epidemiology addressing the unknown causes of non-infectious diseases went back to rely, as for instance Snow had done before the microbiological era, pre-eminently on direct observations in human groups.

The new epidemiology

Individual studies on cancer, non-rheumatic cardiovascular diseases, and psychiatric disease can be traced well back in time, but one can take as a convenient turning point for the rise of the new epidemiology the period around the Second World War. A major stream in the development of the new epidemiology is what could be labelled as the 'tobacco and health story'. Initial observations were either of a statistical nature or of a clinical one. Among the first, one can mention a short and remarkable paper by Pearl (1938) in which he used insurance data to show that the life expectation of smokers was substantially reduced than that of non-smokers. Among the second, one might single out the observation by Ochsner and DeBakey (1939) of the high frequency of smokers among the lung cancer patients coming to their hospital in those early days of thoracic surgery. These were followed by still other statistical findings of a general nature pointing to a dramatic increase of lung cancer rates in males throughout the 1940s. The investigation clearly showing an association 'most probably' (at that time) causal between tobacco smoking and lung cancer was carried out by Doll and Hill (1950). The paper reporting it remains a classic in epidemiology. It is interesting that when they set forth to investigate the etiology of lung cancer, which had become a common disease in the United Kingdom, they were thinking of air pollution, which was at that time very severe in London (the 'London Fog') as an even more plausible candidate than tobacco smoke. As it turned out, the results of their study neatly caused the role of tobacco smoke to emerge, with that of air pollution much less evident.

Many other investigations, starting from the case-control study of Wynder and Graham (1950), contemporary to that of Doll and Hill, confirmed and greatly expanded the original findings. Doll and Hill themselves added another

well-known investigation, following a cohort of British doctors prospectively and reporting the ten years' follow-up results in 1964 (Doll and Hill 1964). This cohort, which has now been followed for fifty years (Doll et al. 2004), provided strong support for the etiological role of tobacco, not only in lung cancer but in a spectrum of neoplastic and non-neoplastic diseases. It is interesting to note (Saracci 1995) that the survival curves from age 35 to age 100 for the 6,813 US-insured subjects observed by Pearl in the late thirties and for the cohort of 34,439 British doctors followed from 1951 to 1971 by Doll and his co-workers show a closely similar loss in median survival (4.9 and 5 years, respectively) when lifetime smokers are compared to lifetime non-smokers (however, with the fifty years' prolongation of the follow-up till 2001, the loss increased to 10 years in the British cohort).

One could safely state that, by 1964, the date of publication of the results of the cohort study by Doll and Hill and that of the first US Surgeon General's report on smoking and health (the 'Terry Report', from the name of the Surgeon General; US Department of Health, Education, and Welfare 1964), the role of smoking in the causation of a number of lethal diseases can be regarded as soundly established. The Terry Report is an extremely valuable document, which can still be read with profit by anyone interested in assessing large amounts of disparate data bearing on an etiological problem. It is interesting from the methodological viewpoint to know that the inability in the initial periods of study to reproduce neoplasms in animals by exposing them to tobacco smoke was regarded—in line with Koch's third criterion—as an important element for questioning the validity of the conclusions drawn from the epidemiological studies. It is certainly not coincidental that at about that time, epidemiologists felt compelled to rethink the criteria to be used to infer causality in general (rather than in infectious diseases) and that A. B. Hill produced a set of guidelines, which can still be used as a reference (Hill 1965). As already noted, this changes the requirement of an agent to be able to reproduce the disease in animals to the more general and optional requirement of biological plausibility. This perspective on causality had the unintended consequence of downplaying almost completely in some circumstances and epidemiological circles the value of animal experiments (incidentally, it can be noted that, using better experimental set-ups, tobacco smoke has produced cancers in experimental animals): this result was an unfortunate one and in flat contradiction to all thinking and practice in biology.

This phase of the development of epidemiology received a new impulse from the two-way exchange between epidemiology and clinical medicine, an exchange which has been a constant feature in the history of epidemiology, as previously exemplified by Ramazzini, a clinician, and Snow, a physician

(pioneering anaesthesia), who had enlarged clinical observations by looking for causes of disease at the population level; conversely, Louis had brought the methods of population studies into the clinical domain to evaluate the effects of medical acts in patients' populations. The 'new epidemiology' clearly highlights this dual exchange. In Great Britain, John Ryle (1899–1950), professor of medicine at Cambridge, moved from the clinic to become the first director of the Institute of Social Medicine in Oxford. The institute had been established at the time of the Second World War to 'investigate the influence of social, genetic, environmental, and domestic factors on the incidence of human diseases and morbidity', and Ryle was a key inspirer of the work of the post-war generation of British epidemiologists who made crucial contributions to the identification of causal factors of chronic diseases.

In a parallel and opposite move, epidemiologic methods were showing their value for clinical research and were increasingly incorporated into a growing stream of 'clinical epidemiology', namely, studies of diagnostic, prognostic, therapeutic, and rehabilitative procedures in populations of sick subjects. A yardstick in this development was the publication in 1972 by Archie Cochrane (1909–88) of a brilliant essay advocating a systematic use of the randomized trial method to evaluate procedures in the clinical and health services areas (Cochrane 1972).

Today: epidemiology in the making

Today, as yesterday, epidemiology as a population approach to health and disease embraces two bodies of knowledge: epidemiological methods of investigation (which are part of scientific methodology) and epidemiological substantive notions developed by the application of such methods (these notions become part of medicine in its biological and social facets).

Both bodies of knowledge have undergone substantial expansion since the Second World War with the development of the 'new epidemiology'. A simple comparison of a mid 1950s edition with a recent edition of any one of the classical textbooks of medicine (e.g. *Cecil Textbook of Medicine,* ninth edition (Cecil and Loeb 1955) versus the twenty-second edition (Goldman and Ausiello 2004) shows how epidemiology has contributed to change and increased, sometimes dramatically, our understanding of the time and space evolution, etiology, and opportunities for control, preventive or therapeutic, of major classes of diseases such as, for instance, ischaemic heart disease, chronic obstructive lung disease, asthma, cancers at several sites, etc., not to mention newer entities like toxic shock syndrome, AIDS, Legionnaires' disease, or Helicobacter infections. This progress has been matched by the emergence, de facto or formally recognized, of a wide spectrum of subspecialties (cancer

epidemiology, paediatric epidemiology, genetic epidemiology, clinical epidemiology, etc.) within epidemiology itself.

On the methodological side, it is sufficient to remember that no text specifically devoted to epidemiological methods was available before 1960, when the book by MacMahon and co-workers (1960) was published, whereas methodology today provides enough matter for ten, eleven, or twenty major books, at different degrees of completeness and complexity.

Study design methods and statistical methods of analysis have been developed in and from the context of problems in epidemiology, rather than by borrowing them from other areas of applied methodology and statistics; in addition, in recent years, a unified approach to the analysis of occurrence data (incidence and mortality) which has also implications for the study design has been developed based on the unifying principle of likelihood inference. These developments have taken place concurrently with an accelerated evolution in the whole field of biology and health, where one can point out four traits, particularly salient in Europe and other economically developed areas of the world:

- The unprecedented advances of research in some domains, fundamental to all other fields of biology and medicine, like immunology, the neurosciences, and, most prominent, molecular and cell biology and genetics. Clearly, switching the study of higher organisms, including humans, from the anatomy and physiology of the phenotype, as has been the case until now, to the direct study of the anatomy and physiology of the genotype opens an entirely new perspective, the full implications of which (preventive, therapeutic, and ethical) are not yet distinctly perceptible.
- The advances in clinical medicine at the diagnostic and therapeutic levels. Until sixty or seventy years ago, effective treatments could be counted on the figures of one hand, so that the only way open—and in many cases effective as well—for disease control was prevention. Nowadays, treatments capable of effectively influencing the length and quality of survival are available for a number of serious conditions, infectious and non-infectious. As a consequence, the balance and the competition between the preventive and curative approach needs to be seen in a fresh light and critically reassessed. A relevant example is mortality from ischaemic heart disease, the marked decrease of which in several western countries appears to be in part due to decreased incidence and in part due to decreased lethality because of better treatment opportunities.
- The escalating costs of all health care delivery systems, whether private, public, or mixed, have brought to the forefront issues of effective and efficient

use of available resources which were of negligible importance or almost unknown fifty years ago.

- The renewed awareness among professionals and the general public of the dependence of health (of humans as well as of other living organisms) on the environment, material and social, personal, local, or general. In parallel with this goes the realization that tangible deterioration of the environment does take place because of short-sighted human activities.

These developments and their interrelationship change the pattern of the factors capable of promoting, damaging, or restoring health and impose a virtually continuous reappraisal and adaptation of the health care system and, more generally, of all plans of action aimed at influencing health.

Tomorrow's horizon

Four major challenges stand on the horizon of epidemiology (which means of epidemiologists) in the coming decades, entailing, as with any challenge, both opportunities and risks.

The challenge of evolving biology

Few would disagree with the following comment by Sir Richard Doll (1993):

> Classical methods of epidemiological research are proving less and less productive as the simple problems are being successfully solved… without some brilliant new inspirations, the rate of discovery of new facts of any importance by the use of these classical methods must be expected to slow down.

A major avenue to maintain momentum is certainly the now pervasive incorporation of concepts and techniques evolving at an impressive pace from such basic disciplines as immunology, molecular and cell biology, and genetics. Biomarkers of exposure, of early effects, and of susceptibility, acquired or innate (genetic), are sharpening the power of all kind of epidemiological studies. In particular, pathogenesis studies investigating mechanisms through which different factors cooperate in producing a disease can to some extent help to identify specific etiological agents in the environment. For instance, if an investigation combining epidemiology with immunology and biochemistry isolates a specific air pollutant as responsible for the induction of asthma attacks, control measures can be addressed specifically to the sources of that pollutant (a more generic approach to all pollutants in the air may simply not be possible). Pathogenesis studies, however, are per se of major interest as they potentially lead to disease control via pharmacological interventions on host factors. It would, however, be wrong for epidemiologists to concentrate in the future almost exclusively on these scientifically highly attractive studies, as this

would imply reducing the investigation of controllable disease determinants in the human environment.

The challenge of evolving environment

While localized environmental health hazards from polluting toxic agents are widespread and have a substantial impact on health, a larger scale and potentially irreversible class of environmental health hazards are emerging. The realization of these hazards stems from the gradual building up and consolidation of knowledge on the size of global environmental changes and on the role of ever-increasing human activities in the production of these changes. Examples include climate change, freshwater shortages, loss of biodiversity (with consequent changes in the functioning of ecosystems), and exhaustion of fisheries (McMichael et al. 2008). These changes are unprecedented in scale, and assessing the resultant risks to populations in the five continents demands innovative approaches, based on empirical data and on alternative scenarios, to modelling environment—disease relationships and evaluating their causal nature.

The challenge of evolving society

Society in most developed countries, and particularly in Europe, is characterized by an ageing demography, reproduction rates below the population replacement rates, a flow of immigration from less developed countries and which is likely to continue in the coming decades, and persistent inequalities in health conditions between different sections of society, in particular with respect to gender, race, and occupational and socioeconomic categories. Monitoring trends and identifying causal factors in this area of 'social epidemiology' has been a long-standing concern of epidemiologists and has received a renewed impulse in the last ten to fifteen years, thus leading to an increasing volume of studies at country and international level. This positive development characterizes epidemiology as a science for justice in health (Saracci 2007) through the epidemiologists' involvement in public health, both from the scientific viewpoint and, when required, from the viewpoint of campaigning for health.

The challenge of diversification versus integration

As in all other scientific and technical branches of activity, epidemiology has recently been, and still is, diversifying and specializing along different axes. A first axis is methodology versus substantive studies; areas of current and future development in methodology include, for instance, methods in genetic epidemiology, modelling of exposure response relationships with multiple longitudinal measurements, and treatment of exposure measurements and errors of measurement in order to reduce misclassification and improve study power. A second axis is the diversification of different fields of substantive interest, for instance,

cancer epidemiology, the epidemiology of ageing, etc. A whole area of specialization is 'clinical epidemiology', which is the application of epidemiological methods within the clinical domain, both for studies evaluating diagnostic, prognostic, therapeutic, and rehabilitative procedures and for evolving formal methods of optimal clinical decision-making. Stemming from it is the rapidly growing branch of 'evidence-based medicine' (EBM), which employs formal methods to assemble and evaluate existing evidence on the effects of medical interventions.

A final axis of diversification tends to separate those who specialize in investigative aspects for routine or research purposes from those who plan and implement interventions; in clinical medicine, this has produced a variety of specialists in purely diagnostic activities (clinical chemists, clinical pathologists, diagnostic radiologists, and imaging specialists); such specialists are different from the therapist, who decides and acts on the basis of the diagnosis; similarly, in the public health area, the epidemiologist may become more and more a pure specialist in etiological and evaluative investigations, leaving to others to decide what to do.

This trend raises three major issues: to what extent can a global view be preserved jointly with specialized or 'subspecialized' technical skills; to what extent, side by side with epidemiologists specializing in different areas, can the figure of the generalist be maintained (again, the analogy with clinical medicine is pertinent, where one of the most difficult present-day problems is the survival and the role of the general physician or the internist); and how can epidemiologists best cooperate in teams of specialists, including groups in charge of taking public health decisions.

Conclusion

Confronting these challenges to the point of overcoming the multiple obstacles to an effective translation of research, thus resulting in health benefits for all people, demands proactive involvement by epidemiologists. Such involvement may take various forms, from assistance to participation in decision-making, and from critique at a social level to frank advocacy actions. It is an engagement of an essentially political inspiration that many epidemiologists may see as outside their exclusive commitment to research, particularly at a time when the already mentioned fast advances in biology propel research into new territories. If this attitude were to prevail, our 'modern' epidemiology would be better labelled 'post-modern', as postmodernism refrains from any utopian project, including the one of universal human betterment (in health as in many other aspects) born with the Enlightenment. It behoves epidemiologists to show that this is not a utopia but the concrete guide of their daily work.

Acknowledgements

Text extract reproduced with permission from Virchow, C. E. A. 'Mortality statistics from the International Congress on Statistics in Paris', in L. T. Rather, ed., *Notes to Part 3: Vital Statistics (Birth, Death and Morbidity Rates): Collected Essays on Public Health and Epidemiology*, Volume 1. Canton, MA. Copyright © 1985 Science History Publications.

Bibliography

Books

Holland, W. W., Olsen, J., and du V. Florey, C., eds. (2007) *The Development of Modern Epidemiology: Personal Reports from Those Who Were There*. Oxford: Oxford University Press. A multi-authored book on the evolution of epidemiology in the second half of the twentieth century. The reports of the 'witnesses' cover extensively the overall picture, specific disease areas, applications and role of epidemiology in related domains, methodology, and historical profiles of regions and countries.

Morabia, A., ed. (2004) *A History of Epidemiologic Methods and Concepts*. Basel: Birkhauser Verlag. A multi-authored book including articles on the history of various aspects of epidemiology as well as ample commentaries on classic papers in epidemiology. The first section (by the book editor) provides an outline, rich in original quotations, of the emergence and evolution of epidemiologic concepts and methods.

Porter, R. (1997) *The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present*. London: Fontana Press. A history of health and medicine, seen both in its internal development and in its relationship to society, as a background to the history of epidemiology.

Rosen, G. (1993) *A History of Public Health* (expanded edn). Baltimore, MD: Johns Hopkins University Press. A standard reference, comprehensive and highly readable.

Collections of articles and essays

***American Journal of Epidemiology*, vols 141 and 142** (1995). On the occasion of its 75th anniversary, the journal reprinted a number of articles published from the late thirties to the late seventies and regarded as of 'historical' relevance. Each is accompanied by a short commentary.

Buck, C. et al. (1988) *The Challenge of Epidemiology: Issues and Selected Readings*. Washington, DC: Pan American Health Organization. An indispensable collection of papers from Hippocrates to the present day.

Greenland, S., ed. (1987) *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods*. Chestnut Hill, MA: Epidemiology Resources Inc. A collection of methodological papers published between 1945 and 1977, covering issues of causal inference and developments in theory and quantitative methods. Each paper is accompanied by a commentary placing it in historical perspective.

Lilienfield, A. M., ed. (1980) *Times, Places and Persons: Aspects of the History of Epidemiology*. Baltimore, MD: Johns Hopkins University Press. A series of essays, most of them documented, by historians and epidemiologists; topics range from numerical methods used in the 1830s to the history of the eradication of smallpox.

US Department of Health, Education, and Welfare. (1964) *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Public Health Service Publication 1103. Washington, DC: US Public Health Service, Office of the Surgeon General. A landmark report that establishes the etiologic role of tobacco smoking in a number of diseases through a rigorous examination of the evidence, mostly epidemiological. Still instructive from a methodological angle.

Websites

Morabia, A. and Vandenbroucke, J. P. (2013) *People's Epidemiologic Library*. <http://www.epidemiology.ch/history/PeopleEpidemiologyLibrary.html>, accessed 8 December 2014. Includes links to other websites on the history of epidemiology.

UCLA Department of Epidemiology, School of Public Health. (2014) *John Snow*. <http://www.ph.ucla.edu/epi/snow.html>, accessed 8 December 2014.

Additional references

Adami, H. O. and Trichopoulos, D. (1999) Epidemiology, medicine and public health. *International Journal of Epidemiology*, **28**: S1005–8.

Boffetta, P. et al. (1998) Multicenter case-control study of exposure to environmental tobacco smoke and lung cancer in Europe. *Journal of the National Cancer Institute*, **90**: 1440–50.

Carmichael, A. G. (1986) *Plague and the Poor in Renaissance Florence*. Cambridge: Cambridge University Press.

Cecil, R. L. and Loeb, R. F.(1955) *A Textbook of Medicine* (9th edn). New York: WB Saunders Company.

Cochrane, A. L.(1972) *Effectiveness and Efficiency. Random Reflections on Health Services*. London: The Nuffield Provincial Hospitals Trust.

Doll, R. (1993) Lecture at the European Educational Programme in Epidemiology 6th Residential Summer Course in Epidemiology, Studium Centre, Florence.

Doll, R. and Hill, A. B. (1950) Smoking and carcinoma of the lung: preliminary report. *British Medical Journal*, **2**: 739–48.

Doll, R. and Hill, A. B. (1964) Mortality in relation to smoking: ten years' observation of British doctors. *British Medical Journal*, **2**: 1399–1410 and 1460–7.

Doll, R., Peto, R., Boreham, J., and Sutherland, I. (2004) Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal*, **328**: 1519.

Dupaquier, J. and Dupaquier, M. (1985) *Histoire de la demographie*. Paris: Perrin.

Goldman, L. and Ausiello, D. (2004) *Cecil Textbook of Medicine* (22nd edn). Philadelphia, PA: Elsevier Health Sciences.

Hackshaw, A. K., Law, M. R., and Wald, N. J. (1997) The accumulated evidence on lung cancer and environmental tobacco smoke. *British Medical Journal*, **315**: 980–8.

Hill, A. B. (1965) The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine*, **58**: 295–300.

Hippocrates. (1923) *Ancient Medicine. Airs, Waters, Places. Epidemics 1 and 3. The Oath. Precepts. Nutriment* (tr. by **W. H. S. Jones**). Loeb Classical Library 147. Cambridge, MA: Harvard University Press.

Holland, W. W. (1999) What should be the concerns of epidemiology? *International Journal of Epidemiology*, **28**: S1009–11.

Hunter, D. J. (1999) The future of molecular epidemiology. *International Journal of Epidemiology*, **28**: S1012–14.

Koch, R. (1893) Ueber der augenblicklichen Stand der bakteriologischen Cholera diagnose. *Zeitschrift für Hygiene und Infektionskrankheiten*, **14**: 319–38.

Lebowitz, J. O. (1970) *The History of Coronary Heart Disease*. London: Wellcome Institute of the History of Medicine.

MacMahon B., Pugh T.F., and Ipsen J. (1960) *Epidemiologic Methods*. Boston, MA: Little, Brown and Company.

Maggi, L. (2000) Bearing witness for tobacco. *Journal of Public Health Policy*, **21**: 296–302.

McMichael, A. J., Friel, S., Nyong, A., and C. Corvalan. (2008) Global environmental change and health: impacts, inequalities, and the health sector. *British Medical Journal*, **336**: 191–4.

Ong, E. K. and Glantz, S. A. (2000) Tobacco industry efforts subverting International Agency for Research on Cancer's second-hand smoke study. *Lancet*, **355**: 1253–9.

Ochsner, M. and DeBakey, M. (1939) Primary pulmonary malignancy. Treatment by total pneumonectomy; analyses of 79 collected cases and presentation of 7 personal cases. *Journal of the American College of Surgeons*, **68**: 435–51.

Pearl, R. (1938) Tobacco smoking and longevity. *Science*, **87**: 216–17.

Pearl, V., Worden, B., and Lloyed-Jones, H., eds. (1981) *Essays in Honour of H. R. Trevor-Roper*. London: Duckworth.

Pearce, N. (1999) Epidemiology as a population science. *International Journal of Epidemiology*, **28**: S1015–18.

Proctor, R. N. (1999) *The Nazi War on Cancer*. Princeton, NJ: Princeton University Press.

Ramazzini, B. (1982) *De morbis artificum diatriba* (Italian tr. by **I. Romano**, **V. Romano**, **and F. Carnevale**). Florence: La Nuova Italia Scientifica.

Reid, D. D. (1948) Sickness and stress in operational flying. *British Journal of Social Medicine*, **2**: 123–31.

Riunione degli scienziati italiani. (1939) *Atti della prima riunione degli scienziati italiani* (4th edn). Pisa: Nistri-Lischi.

Saracci, R. (1995) Smoking and death. *British Medical Journal*, **310**: 600 and 672.

Saracci, R. (1999) Epidemiology in progress: thoughts, tensions and targets. *International Journal of Epidemiology*, **28**: S997–9.

Saracci, R. (1999a) Discussion. *International Journal of Epidemiology*, **28**: S1023.

Saracci, R. (1999b) Epilogue. *International Journal of Epidemiology*, **28**: S1024.

Saracci, R. (2007) Epidemiology: a science for justice in health. *International Journal of Epidemiology*, **36**, 265–8.

Saracci, R. (2010) 'Introducing the history of epidemiology', in **W.W. Holland**, **J. Olsen**, **and C. du V. Florey**, eds, *The Development of Modern Epidemiology: Personal Reports from Those Who Were There* (3rd edn). Oxford: Oxford University Press, pp. 31–40.

Saracci, R. (2011) 150 years of epidemiology in united Italy: notes for a history yet to be written. *Epidemiologia e Prevenzione*, **35** Suppl. **2**: 14–7.

Singer, C. and Underwood, E. A. (1962) *A Short History of Medicine* (2nd edn). Oxford: Clarendon.

Susser, M. (1999) Should the epidemiologist be a social scientist or a molecular biologist? *International Journal of Epidemiology*, **25**: S1019–22.

Virchow, C. E. A. (1985) 'Mortality statistics from the International Congress on Statistics in Paris', in **L. T. Rather**, ed., *Notes to Part 3: Vital Statistics (Birth, Death and Morbidity Rates): Collected Essays on Public Health and Epidemiology*, vol. 1. Canton, MA: Science History Publications, pp. 589–93.

Wall, S. (1999) Epidemiology in transition. *International Journal of Epidemiology*, **28**: S1000–4.

Winkelstein, W. Jr. (1995) A new perspective on John Snow's communicable disease theory. *American Journal of Epidemiology*, **142**, S3–9.

Winslow, C. E. A. (1980) *The Conquest of Epidemic Disease*. Madison, WI: The University of Wisconsin Press.

Wynder, E. L. and Graham, E. A. (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma: a study of 684 proved cases. *Journal of the American Medical Association*, **143**: 329–36.

Chapter 2

Important concepts in epidemiology

Olli S. Miettinen

Introduction to important concepts in epidemiology

For my writing about important concepts in epidemiology—about how I see them—the context now is quite different from what it was when I wrote about them for this book's most recent previous edition, for I have invested a lot of thought in these concepts during the production of two recent books of mine (Miettinen 2011; Miettinen and Karp 2012). So, upon having formulated the importance hierarchy of the concepts relevant to this book, I largely draw my definitions and explications of those concepts from these two sources.

Epidemiology vs epidemiological research

The most centrally relevant, and hence the most important, concept here obviously is that of *epidemiology*. To me, *epidemiology* has a singular meaning: it is (the practice of) community medicine (Miettinen 2011; Miettinen and Karp 2012), although to most teachers of the subject the concept of epidemiology has 'study' as its proximate genus. Moreover, the specific difference unique to epidemiology within this genus is commonly said to be the study's focus on something like 'distribution and determinants' of 'disease' in populations. I edit those definitions to mean that epidemiology is taken to be a line of research—specifically, research on the (rates of) occurrence of phenomena of human health (Miettinen and Karp 2012).

The International Epidemiological Association (IEA) defines *epidemiology* as 'the study of the occurrence and distribution of health-related states or events in human populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems' (Porta et al. 2008).

I now distinguish, more sharply than ever before, between *epidemiology* and *epidemiological research*; and my concept of the latter now is very different from those above. The concept of epidemiological research naturally is to be deduced from that of epidemiology, whereas epidemiology is defined without any reference to research. In this framework, I define *epidemiological research* as 'research—*any* research, however "basic"—aimed at advancement of the practice of community medicine' (Miettinen and Karp 2012).

In these terms, this book is not about teaching epidemiology; it is about teaching epidemiological research. And this book is about only one of the two principal segments of epidemiological research; it is about teaching population-level rather than laboratory-based epidemiological research—research that addresses the occurrence of illness—that is, morbidity (Miettinen 2011)—in human populations. This book is about teaching research on morbidity in human populations.

The *IEA Dictionary of Epidemiology* (Porta et al. 2008) defines *morbidity* as any departure, subjective or objective, from a state of physiological or psychological well-being. In this sense *sickness*, *illness*, and *morbid condition* are similarly defined and synonymous (but see *disease*). The WHO Expert Committee on Health Statistics noted in its sixth report (1959) that morbidity could be measured in terms of three units: (1) persons who were ill, (2) the illnesses (periods or spells of illness) that these persons experienced, and (3) the duration (days, weeks, etc.) of these illnesses. In addition, it defines *morbidity rate* as 'a term, preferably avoided, used to refer to the incidence rate and sometimes (incorrectly) to the prevalence of disease'.

Rates of morbidity

Intimately related to my concept of morbidity (which is very different from that of the IEA) are those concepts concerning the frequency of occurrence of illness in human populations. The first concept is the *rate of incidence*, having to do with event-type phenomena such as complications of delivery or vaccination and diagnosis (rule-in) about a particular type of cancer. The incidence rates of these complication-type phenomena, or of neonatal death, for example, are in the form of a *proportion*. For diagnosis (rule-in) about a cancer, the incidence rate generally is of the form of a *density*: the number of the events occurring in a unit amount of population-time. From such a rate of incidence can be derived the *cumulative* incidence rate for a span of time. This too is a proportion, inherently conditional on not succumbing to an extraneous cause of death in the time period at issue. In addition, there are *rates of prevalence* for state-type phenomena such as congenital malformations. All rates of prevalence are in the form of *proportions*.

These elementary concepts of rates of morbidity arguably are the most important ones in both epidemiology and epidemiological research, as they are the central concern in both. And they deserve attention here for the added reason that an *anomalous conception of rates* has had a high rate of prevalence among teachers of epidemiological research ever since its introduction four decades ago (Elandt-Johnson 1975). This anomaly is the notion that a proportion cannot properly be thought of and referred to as a rate—that a rate inherently has a temporal dimension, so that only *incidence density* among the quantities listed above is a rate.

The IEA dictionary says under 'Rate' that, in epidemiology, rate is 'an expression of the frequency with which an event [*sic*] occurs in a defined population, usually in a specified period of time. . . . the term *prevalence rate* is to be avoided, because prevalence cannot (and does not need to) be expressed as a change in time . . . In contrast, the force of mortality and the force of morbidity (hazard rate) are proper rates, for they can be expressed as the number of cases developing per unit time divided by the total size of the population at risk.' For 'Cumulative incidence, cumulative incidence rate (syn: incidence proportion, average risk)' the IEA definition is 'the number or proportion of a group (cohort) of people who experience the onset of a health-related event during a specified time interval; this interval is generally the same for all members of the group, but, as in lifetime incidence, it may vary from person to person without reference to age'. Under 'Prevalence' is said that 'it is a proportion, not a rate.'

The IEA dictionary thus displays considerable confusion about epidemiological rates and is inconsistent in what it says about them. It does not reflect an understanding of what incidence density and cumulative incidence are, and even the term *incidence density* (which I introduced in 1976, along with *cumulative incidence*) remains alien to its authors, although it is commonplace in epidemiological research. In addition, the IEA dictionary propagates the notion that proportion-type measures of the frequency of the occurrence of an illness are not rates.

The essence of the anomaly in Elandt-Johnson's conception of rates (Elandt-Johnson 1975) is that it represents a bizarre revolt against the English language, as this language is well established in general usage, in various scholarly contexts, in the professional realm of epidemiology, and in epidemiological research à la Richard Doll, Abraham Lilienfeld, Brian MacMahon, etc. My Concise Oxford English Dictionary (Pearsall 1988) says that *rate* is 'a measure, quantity, or frequency, typically one measured against some other quantity or measure'. And my venerable American Heritage dictionary (Morris 1969) specifies one of the three meanings of *rate* as 'a quantitative measure of a part to a whole; proportion: *The birth rate*; *a tax rate*'.

Populations, open vs closed

Although very swift to adopt Elandt-Johnson's major misconception about epidemiological rates, teachers of epidemiological research have been very slow to grasp the fundamental duality in *types of population* (described in Miettinen 1985). This duality derives from some populations being open to exit while others are closed to exit, as some populations' membership is defined by a state, for the duration of that state, whereas membership in others is defined by an event, for the rest of time (eternally) thereafter. Linguistically, the distinction is between *dynamic populations* (having turnover of membership) and *cohorts*. To many teachers of epidemiological research at present, *cohort* seems to be a synonym for 'population', with some cohorts 'fixed' and others 'dynamic'.

The IEA dictionary defines *cohort* as '1. The component of the population born during a particular period and identified by period of birth so that its characteristics (e.g., causes of death and numbers still living) can be ascertained as it enters successive time and age periods. 2. The term "cohort" has been broadened to describe any designated group of persons who are followed or traced over a period of time, as in cohort study (prospective study).' The term *dynamic population* it defines as 'a population that gains and loses members; all natural populations are dynamic—a fact recognized by the term *population dynamics*, which is used by demographers to denote changing composition'. The term *fixed cohort* it defines as 'a cohort in which no additional membership is allowed—that is, it is fixed by being present at some defining event ("zero time"); an example is the cohort comprising survivors of the atomic bomb exploded at Hiroshima'. There is no definition given for *dynamic cohort*.

Etiology, etiological study

Epidemiological research on morbidity is principally about *etiology*; but the concept of this—'causal origin, or causal explanation' (Miettinen 1985)—remains quite commonly not understood by teachers of epidemiological research. To underscore the retrospective nature of this genre of causality in medicine, I now prefer my neologism *etiogenesis* (Miettinen 2011).

The IEA dictionary defines *etiology* as 'literally, the science of causes, causality; in common usage, cause'. But there is no such science; and tautology is not literally, nor otherwise, the science of unnecessary repetition.

From understanding the inherently retrospective nature of the etiology/etiogenesis concept—from the vantage of a case or a series of cases of an illness backward to is causal origin—flows understanding of the *central role of a case series* in any study of the etiogenesis of an illness. This means understanding the need to build the study around a series of cases of the illness, a series in which

some of the person-moments are associated with a positive history for the (potentially) etiogenetic factor, others with a positive history for the alternative to this in the causal contrast. Critical in this construction is the understanding that any case series has meaning in epidemiology and in epidemiological research only as a source of the numerator inputs to the calculation of rates (or quasi-rates) of occurrence of the cases in a defined referent population-time (or series of person-moments) for the rates. From these elementary understandings flows the concept of *the etiogenetic study*, in which the case series is coupled with a base series, a fair sample of the study base (Miettinen and Karp 2012).

Before these rather recent elementary understandings of etiogenetic research, A. B. Hill—the statistician with a major role in the development of randomized trials to address the other genre of causal problems in medicine—posited the highly influential but seriously misguided idea that this research, while non-experimental, should be designed with the intervention experiment paradigmatically in mind (Hill 1953). This led to the concept of the 'prospective' study in etiologic/etiogenetic research; but because of its problems of practicability, 'retrospective' study remained commonplace as well (US Department of Health, Education, and Welfare 1964). These two types of study were subsequently renamed *cohort* study and *case-control* study. Correction of the respective fallacies in these leads to the singular essence of the etiogenetic study sketched above (Miettinen and Karp 2012).

Regrettably, and incomprehensibly to me, 'cohort' and 'case-control' studies remain de rigueur in the prevailing culture of etiologic/etiogenetic research and in the teaching of this, just the same. In line with this, the IEA dictionary (Porta et al. 2008) defines them without any critical notes attached. It does not point out that the 'cohort' study does not have the case series with base series structure, with etiogenetic histories as of these person-moments, and that, while a 'case-control' study is one 'of persons with the disease (or another outcome variable) of interest and a suitable control group of persons without the disease . . . comparing the diseased and nondiseased subjects with regard to how frequently the factor or attribute is present . . . in each of the groups (diseased and nondiseased)', it does not have even a semblance of a study base, with the causal contrast in it and with case series and base series from it.

Confounding

As at issue here is non-experimental research on causality; a very important concept in it is *confounding*. In reference to studies of the etiogenetic type of causation, the core idea about confounding is this: a health event's association with a potentially causal antecedent (in an unbiased study base) is a reflection

of causality, confounding, or both. In this, the antecedent involves the causal–contrast duality of histories (cf. 'Populations: open vs closed'), and any potential confounder is an extraneous determinant of the rate of occurrence of the illness. A potential confounder is an actual confounder (of the study base) if it has a differential distribution between the index and reference segments of the study base (representing positive histories for the factor and for its alternative, respectively).

The IEA dictionary (Porta et al. 2008) says under 'Confounding bias (syn: confounding)' that this is '1. Bias of the estimated effect of an exposure on an outcome due to presence of a common cause of the exposure and the outcome . . . 2. Bias of the estimated effect of an exposure on an outcome due to baseline differences among exposure groups in the risk factors for the outcome'. However, both of these are misconceptions of confounding (of the study base) in etiogenetic studies. Age and gender are the two archetypical confounders in etiogenetic studies, and neither one of these can be thought of in causal terms, as neither admits a causal contrast due to the inherent absence, at a given person-moment, of an alternative to a person's actual age or gender. In truth, both of the associations definitional to a confounder (above) are acausal (meaning that causality is not an issue). And 'baseline differences among exposure groups', insofar as they have anything to do with an etiogenetic study, have to do with the study's source population when it is of the cohort type rather than the dynamic type. But even with a cohort as the source population, the study population (forming the population-time of the study base) is dynamic, as membership in it is defined by transient states (the state of being alive as one of these). Thus, differences between the index and reference aggregates of population-time are relevant (cf. 'Etiology, and etiological study').

The teacher's challenge

In closing, I harken back to the preface of a recent introduction to epidemiological research (Miettinen and Karp 2012); there, Karp and I commented on how challenging it has been to us, and to others too, to come to grips with the concepts of this research, and how much effort the attainment of understanding of them thus requires of students. Here I add that *teaching* of the research also requires *much effort*, notably in preparation for it. For, as is evident from the foregoing, an obviously challenging aspect of the preparation is steering around the authoritarian but malformed versions of even the most important concepts of this research in the IEA dictionary (Porta et al. 2008). The IEA would do well to arrange public discussions of important concepts of epidemiological research instead of periodically consolidating misconceptions about these concepts.

Acknowledgements

Text extracts reproduced from Porta, M. et al. *A Dictionary of Epidemiology. A Handbook Sponsored by the IEA*. Fifth edition, Oxford University Press, Ney York, USA, Copyright © 2008, by permission of Oxford University Press, USA.

References

Elandt-Johnson, R. C. (1975) Definition of rates: some remarks on their use and misuse. *American Journal of Epidemiology*, **202**: 267–71.

Hill, A. B. (1953) Observation and experiment. *New England Journal of Medicine*, **248**: 995–1001.

Miettinen, O. S. (1985) *Theoretical Epidemiology. Principles of Occurrence Research in Medicine*. New York: John Wiley & Sons, Inc.

Miettinen, O. S. (2011) *Epidemiological Research: Terms and Concepts*. Dordrecht: Springer.

Miettinen,O. S. and Karp, I. (2012) *Epidemiological Research: An Introduction*. Dordrecht: Springer.

Morris, W., ed. (1969) *The American Heritage Dictionary of the English Language*. Boston, MA: Houghton Mifflin Company.

Pearsall, J., ed. (1988) *The New Oxford Dictionary of English*. Oxford: Clarendon Press.

Porta, M., ed., Greenland, S., Last, J. M., associate eds. (2008) *A Dictionary of Epidemiology. A Handbook Sponsored by the IEA* (5th edn). Oxford: Oxford University Press.

US Department of Health, Education, and Welfare. (**1964**) *Smoking and Health*: *Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Public Health Service Publication 1103. Washington, DC: US Public Health Service, Office of the Surgeon General.

WHO Expert Committee on Health Statistics. (1959) *World Health Organization Meeting in Geneva*. World Health Organization.

Chapter 3

Study design

Jørn Olsen and Olga Basso

Introduction to study design

In our experience, the most basic epidemiologic concepts and principles are understood by the majority of the students early in the course, and the topic can be made as complex as the skills of the teachers and students permit. At all levels, there are challenging and important issues to discuss. An important aim of teaching, at least at the university level, is to make students aware of what we—and they—don't know, and much in epidemiology is still unknown.

A study is a process—documented throughout from design to execution—that aims at providing empirical evidence on a given issue. An epidemiologic study may start with a public health problem, an hypothesis concerning the etiology of a disease, or may be prompted by the available research opportunities.

Studies may start with a health problem that needs a solution, or because a new, or differently formulated, hypothesis may lead to a better understanding of a phenomenon. Also, a study may be worthwhile if it provides data of better quality for exploring an existing hypothesis. An hypothesis is justified not only by its biological plausibility but also based on its epidemiologic coherence. Does the incidence of the disease follow a pattern that, at least to some extent, correlates with exposure distributions over time and between populations? It is, for example, highly unlikely that use of cell phones is a strong risk factor for brain cancer, at least if the induction time is less than ten years. Some studies may be undertaken without any prior hypothesis, such as genome-wide association studies (see chapter 12).

In any study, careful characterization of the outcome is a fundamental step. For example, if one wishes to study infertility, would that be defined solely on the basis of seeking medical help due to an inability to conceive, or should researchers adopt a wider definition, asking a sample of couples in the general population for how long they have tried to conceive as an indicator of infertility using a given waiting time (e.g. 1+ years) as a cut-off point? The

occurrence of this infertility (no matter how it is defined) as a function of exposure to the putative causal factor is addressed in what Miettinen calls 'the object design'. Avoiding bias is addressed in the 'method design' (Miettinen and Karp 2012).

In general, the 'first' study on a new hypothesis is very important, since subsequent studies may be influenced by its results. For example, when it was first reported that vasectomy may increase the risk of prostate cancer, a number of men who had undergone vasectomy were checked for prostate cancer (either at the patient's request or because the physician was alerted to the problem). As prostate cancer can go undetected for a long time, such a mechanism may artificially produce a higher proportion of diagnosed prostate cancer in men who have been vasectomized than in those who have not. To overcome this problem and achieve equal accuracy of information, only severe cases that would not be missed regardless of the exposure could be considered. Comparable accuracy is particularly important when an outcome is rare and the estimated effect is expressed with a relative measure (i.e. relative risk, hazard ratio, or odds ratio), as poor specificity results in greater bias towards the null than poor sensitivity. This is due to the fact that, with poor specificity, a large number of false positives will 'contaminate' the small number of cases, thus diluting any effect the exposure of interest may have.

The study design should be dictated by its aim(s), if possible. The challenge is to find a design that most effectively and validly addresses the research hypothesis. If the optimal design is not feasible for ethical or practical reasons, it is important to consider if it is worthwhile pursuing the study at all. The optimal sample size may well be zero.

The main aim of the teaching should be to focus on the logical link between the study aim(s) and the design. Key issues in teaching study design are related to how the entities (determinants, outcome, confounders, and modifiers) under study are formulated and whether the results will provide the evidence for which they are designed.

Teachers should try to remember the difficulties they themselves encountered as students of the discipline. This is often a good indicator of what needs particular attention in teaching. Let the students interrupt. Insist on questions from the students at regular intervals and present numerous examples, preferably from published papers. Ask questions to test the level of understanding and invite students to evaluate potential sources of bias that may have been overlooked by the authors. Selecting papers with blatant design flaws is a good way to make students aware that even studies published in high-impact journals may suffer from major design problems; however, the emphasis should be on normal papers.

At the undergraduate level, students should be taught the core concepts of basic epidemiological study design and should be able to accurately describe and characterize such design. For graduate students, teaching should incorporate new and more specialized study designs and methods, as well as creative ways to exploit data that are often available, such as birth registries (e.g. we have learnt much about the familial risk of several reproductive outcomes by studying pregnancies from maternal and paternal relatives of the index cases). New laboratory techniques and advancement in analysing causal inference have also greatly increased the number of options available in the design and analysis of data.

Epidemiology often rests on the observation of occurrence of a disease as a function of its potential determinants. For ethical reasons, these determinants cannot in most cases be manipulated by the researcher. Most research, at least within public health epidemiology, is therefore non-experimental. In clinical epidemiology, however, it is often possible to manipulate treatment in a way that makes scientific inference less vulnerable to bias, as in the randomized controlled trial.

Epidemiologic research aims to capture the disease experience within a population according to lifestyle factors, treatment, environmental exposures, genetic factors, etc. and to make inferences from such experience. Rather than focusing upon dissimilarities, the teacher should focus on the similarities of the various designs. The common aim is to capture the underlying population experience, for different purposes and with different efficiency (and accuracy), while at the same time pursuing the maximization of validity at the lowest possible cost.

An epidemiologic study is usually a simplification of reality (e.g. one outcome, often binary, and a limited number of determinants). Since no study can specify all modifiers and confounders, this is generally acceptable and, possibly, the only way to proceed in many situations; however, it is important to be aware of this limitation. In particular, it is important to remind students that an exposure that may be harmful for a particular outcome may be protective for others, and vice versa. For example, although sun exposure is a known risk factor for skin cancer, it is also the main source of vitamin D, and vitamin D deficiency is a risk factor for several illnesses (Holick 2007).

Theoretical epidemiology, that is, research into the concepts and methods of the discipline, has developed rapidly during the last thirty years (Miettinen 1985a, 1988, 1999; Rothman et al. 2008; Szklo and Nieto 2012) leaving many of the disease- or exposure-oriented epidemiologists behind. This need not cause problems if the teacher sticks to basic and well-defined concepts and principles. Many (perhaps most) of the important epidemiologic studies are performed by

people with limited knowledge of theoretical epidemiology and biostatistics, because the idea (or the hypothesis) is the most important part of the study. The rise and fall of new methods has made it difficult to reach consensus on proper textbooks but many of the older textbooks are outdated, especially concerning the understanding of case-control studies and concepts like confounders, intermediates, and effect measure modifiers. These developments are mostly related to achieving more rigorous ways of collecting the available information and a more precise understanding of the ability to estimate the effect measure of choice in the underlying population (Axelson 1979; Miettinen 1985b, 1999; J. Olsen et al. 2010).

In recent years, genetics has increasingly become part of epidemiologic studies, along with the use of biomarkers of exposure, and not all teachers will be familiar with these techniques. Several designs have been proposed for epidemiologic studies that incorporate genetic analyses, and some of these should be proposed to graduate students (Weinberg and Umbach 1999, 2000; Weinberg 2003; Weinberg and Umbach 2005; Weinberg et al. 2007; Lawlor et al. 2008).

While the fundamental principles of epidemiologic research still apply, investigators with some knowledge of genetics and basic science are better equipped to designing and conducting these studies. Studies in which a large number of genetic polymorphisms are examined will also require statistical techniques capable of meeting some of the challenges posed by numerous statistical comparisons.

Other areas of epidemiology that have recently undergone a profound change are related to causal inference (VanderWeele and Vansteelandt 2009; Valeri and VanderWeele 2013) and to assessing potential confounders, often via the use of directed acyclic graphs (DAGs), to establish which covariates are actually confounders and which should not be included in the model due to their potential for biasing the estimates (Greenland et al. 1999; Hernán et al. 2002; Pearl 2010; Wilcox et al. 2011; Howards et al. 2012).

Teaching objectives

Undergraduate students should be aware of the core measures of association in epidemiology, such as prevalence, incidence, relative risk (or risk ratio), risk difference, incidence rate ratio, odds ratio, rate difference, etc., as well as of the strengths and weaknesses of the basic designs. They should be able to read epidemiological papers critically, whether the publisher is a drug company, a newspaper, or a scientific journal. Students should also be able to read articles that use standard designs and to comment on the relevance of the design in light of the aim of the study. They should be aware of the major potential

sources of bias and be able to identify them in individual studies. They should be able to explain how randomization, blinding, and placebo administration operate in minimizing bias and confounding in randomized controlled trials. They should be aware of ethical principles and principles of right to privacy and data protection. Students should also be aware of how epidemiological designs may be used to monitor health-care interventions, including screening, over time. They should understand and be able to describe the strengths of these designs in comparison with non-systematic reporting of side effects of drugs, accidents, etc.

Students should also understand the meaning of the basic concepts of measuring observation time in epidemiology and know the concept of immortal observation time. They should be able to define epidemiological problems and know the differences between medical sociology, behavioural economics, medical psychology, and epidemiology. It is epidemiologic research to study the health consequences of smoking but, for example, not why people smoke or engage in unsafe sex.

Graduate students in epidemiology should have a more detailed knowledge of the different design options and be able to design typical studies themselves. They should be able to understand the concepts of validity, precision, efficiency, and study power. At the end of the course, they should be able to peer review a paper and write a study protocol using one of the standard designs in epidemiology. They should also follow the debate on policy issues in journals or professional websites (such as <http://www.ieaweb.org>).

Since epidemiologists often try to identify causes of diseases or health conditions, the syllabus should include a session on the concept of causation and counterfactuals, and its consequences for the design and interpretation of the results (Hill 1965; Lewis 1973; Mackie 1975; Rothman 1976; Susser 1991; J. Olsen 1993). DAGs (Pearl 2009) may be useful tools for identifying sources of bias and confounding (Greenland et al. 1999). They could be introduced here and further elaborated in a course on bias and confounding.

Teaching should be based on up-to-date concepts and methods, and the teacher needs to know some of the basic literature in theoretical epidemiology, at least from the last decades (Rothman et al. 2008; J. Olsen et al. 2010; Ahrens and Pigeot 2013). Teachers must be familiar with fundamental concepts, such as bias, confounding, effect modification, and DAGs (Miettinen 1981, 1988; Weinberg 1993; Greenland et al. 1999; Pearl 2010).

Since epidemiology deals with often sensitive data and/or sensitive questions, students should be aware of how to do properly conducted studies (International Epidemiological Association (IEA)/The European Epidemiology Group 2007).

Teaching contents

Rather than simply classifying designs as randomized controlled trials, follow-up studies, case-control studies, cross-sectional studies, or correlation studies, students at all levels should be taught about the key parameters that characterize a given design. These are discussed in the following sections.

The unit of observation

This could be an individual or a group of individuals. When the units of observation consist of an aggregate of individuals, the study is often called an ecological or correlation study. The unit of observation may also be an event, rather than an individual (e.g. a pregnancy). If only one event per individual is analysed, statistical independence is no problem. If, however, some individuals contribute to more than one event, this could make the analysis more complex, as the assumption of independence maybe violated (Hardin and Hilbe 2003; Basso 2007; Howards et al. 2007; J. Olsen, 2008; Howards et al. 2012). The unit of the observation may also be only a part of the body, as when brain cancers are compared for the part of the brain exposed to mobile phone use and compared with the occurrence of brain cancer in unexposed parts of the brain.

Type of population

Populations studied in epidemiology can be of two main types: closed cohorts (membership is defined by a given event, and no exit is possible), and dynamic populations (where a given state, e.g. residence in a given area, defines membership in the cohort, and exit occurs when that state terminates). Closed cohorts are used if certain exposure measurements need to be taken at a certain time period during the life course or if moving in or out of a study area may cause selection bias.

Allocation of the exposure under study

Depending on the type of study, allocation of exposure may be performed by the researcher to learn about the health consequences of a given intervention (experimental design). In cases of potentially harmful exposures, induced by nature, self-selection, or other mechanisms, often the researcher cannot manipulate the exposure but can organize observations in order to learn about the effect of such exposures (non-experimental design).

A 'natural experiment', like an earthquake, is not an experiment in the scientific sense of the word—it is not an exposure which was induced in order to learn about its effect. Although this specification may seem obvious, it is important to make the students properly understand what characterizes an experiment and

what does not. In an epidemiologic experiment the exposure is under the control of the researcher and the aim is to learn about the health consequences of this exposure.

The timing of the observations

Since diseases occur over time, *a longitudinal* recording of exposures and disease is preferable—a sequential set of observations in real time—as in the cohort study and—when possible—in the case-control study. Situations in which exposures, outcomes, and other factors are recorded at the same point in time using prevalence data without attempting to reconstruct the exposure history in relation to the outcome are usually defined as *cross-sectional* studies. These studies often suffer from a number of problems that hamper the interpretation of the findings of associations, largely due to the fact that the correct temporal sequence (cause occurring before the effect) often cannot be established with certainty. 'Reverse causation' may be a problem when the disease impacts lifestyle or metabolism of environmental pollutants. Cross-sectional studies are better suited to provide descriptive data on the occurrence (prevalence) of diseases and exposures but may be of etiologic value in genetic studies or in other situations where the exposure does not change over time.

Definition of the relevant etiologic time window

Based upon the available knowledge of the condition under study and of the hypothesized effect of the exposure, the appropriate etiologic time window must be defined (i.e. cigarette smoking is unlikely to be the only cause of lung cancer in subjects who started smoking six months before being diagnosed, and a potentially teratogenic drug taken in the fourth month of pregnancy cannot be the cause of a cleft palate because the fusion of the palate is over by then). It is important not only to consider the time it takes for a given exposure to initiate the disease process and to identify the relevant window of vulnerability but also to take into consideration the latency time, that is, the time it takes before the disease has reached a stage where it can be diagnosed.

Definition of the study base (or source population)

The study base, or source population, is the population experience over the time period under study—the actual person–time experience that is the basis of the inference (Miettinen 1985a). It is *a primary* study base when the source population is defined before the case series is identified, and the challenge is then to obtain complete ascertainment of the cases. It is *secondary* when the case series is identified first, and the challenge is then to identify the source population that originated *that* case series.

Sampling the study base

When the study base is sampled according to the exposure status, for example, exposed versus unexposed, we usually talk about a *cohort* study. If we first identify cases and then sample from the study base, the study is defined as a case-control study. Since cohort studies are expensive, they often address more than one exposure and can cover the entire population in a region, like the Framingham study.

Types of data

Primary (or ad hoc) data are those collected for the purpose of a specific study, while secondary (or antecedent) data are generally collected for other purposes, for example, registers, medical files, etc.

The terms 'prospective' and 'retrospective' have been used by epidemiologists in many different ways and some believe it is best to avoid these terms altogether. Sometimes, however, exposure data collected prior to the outcome are referred to as prospective, and exposure data collected after the outcome as retrospective. The terms should, however, be avoided, at least with respect to causal inference. Causal inference is always forward in time, regardless of the design. In the case-control study, we reconstruct the occurrence of the putative determinants in the past to learn about its consequences. We do not make inference from the effect to its possible cause. A case-control study is therefore not a retrospective study from the perspective of causal reasoning, and the term 'tro-hoc' study (cohort spelled backwards; Feinstein 1973) bears witness to the misunderstanding of the underlying structure of this design.

It may be useful to illustrate the use of different designs in the evaluation of a hypothesis by means of an example. This particular example illustrates studies that often progress from inexpensive to more expensive and from simple to more complex designs.

> We may, for example, observe that birth weight tends to be higher in areas with a high fish intake (S. F. Olsen 1993).

This hypothesis could, at first, be inexpensively tested with a correlation study, since birth weight data are often available and dietary surveys or sales of food items can provide exposure data. Such a simple study (correlation or ecological study) could show that birth weight tends to be higher in populations characterized by a high fish intake; however, fish intake need not be causally related to birth weight, as we have no information on whether it was the fish-eaters who had the largest babies (ecological fallacy). The association could also be caused by genetic factors, differences in smoking habits, social status, etc. It could, of course, be due to growth-promoting factors in fish, or factors that delay the

onset of labour. This type of study will not in itself provide any effect estimates related to individual risks.

The next step may be to carry out a case-control study, using as cases babies with high birth weights. Controls may be sampled from all pregnant women and we could then try to reconstruct fish consumption during pregnancy. In this study, we would seek information on smoking, social conditions, and other potential determinants of birth weight and fish consumption. Gestational age may be a mediator in this relation, if fish consumption delays labour, and this requires particular attention in the analysis. If such a study showed an association, the evidence would be more convincing; however, we would probably still doubt the quality of the data on dietary habits and we would have no control of confounding by unknown factors.

The next step would be to set up a follow-up study where we recruit pregnant women with a wide variety of levels of fish intake recorded during the entire pregnancy in real time and then compare newborns in this group with the newborns of pregnant women who did not eat fish during pregnancy. The non-exposed group may provide the expected gestational age and birth weight in the exposed cohort, had the exposed cohort not been exposed (counterfactual reasoning: i.e. a given factor is causal if, in its absence, the effect would not have occurred at that point in time (Lewis 1973)). We would attempt to make this group comparable to the exposed group by taking into account potential confounders. In this study we have better information on dietary habits but we would still not fully resolve the problem of potential residual confounding by unknown determinants of foetal growth and pregnancy duration that may also influence fish intake. In fact, such a study might indicate that frequent fish-eaters have smaller babies, perhaps because fish may also be contaminated with pesticides or heavy metals that could impair foetal growth or because some women who eat high amounts of fish during pregnancy do so because they previously experienced an adverse reproductive outcome.

The final logical step would be to set up a randomized controlled trial in which pregnant women would be allocated to high and low intakes of fish, or the presumed active components in fish (*n*-3 fatty acids). If such a trial were large and could be double blinded (which may be possible with *n*-3 fatty acids), it would be unlikely that those who were randomly allocated to receive the active component would differ from those who received the placebo. If the placebo provided a similar energy supply, birth weight should be similar in the two groups if the null hypothesis is true (i.e. *n*-3 fatty acids have no effect except that associated with energy intake). If this study were to show an effect, it would provide a strong case but still no certainty. A greater degree of evidence would be provided by a meta-analysis of all existing eligible trials on the topic. In general,

we only provide evidence that will move personal beliefs towards or away from a causal effect but never to probabilities of 0 or 1. Only the counterfactual comparison, comparing the same people being both exposed and unexposed at the same time, would justify causal inference, and this is not possible.

Hormone-replacement therapy, which appeared to be protective against coronary heart disease in observational studies (Grodstein et al. 1996, 2000; Varas-Lorenzo et al. 2000), was found to be associated with increased risk in the Women's Health Initiative (WHI) randomized trial (The Women's Health Initiative Study Group 1998). In 2008 Hernán et al. reanalysed the data from the Nurses' Health Initiative to try to emulate a randomized trial, which yielded findings similar to those of the WHI trial (Hernán et al. 2008).

Using the fish oil example in teaching will be a sober introduction to the need for being cautious in making causal inference. The randomized controlled trial is, in theory, the most simple to design, analyse, and interpret. For logistic reasons, however, it is often the most difficult to carry out, especially if it is of long duration. It is also important to recognize that randomized trials provide only a degree of guarantee of comparability at baseline if the study is large. Compliance to the allocation of the exposure will almost never be perfect (and far from perfect in a trial of long duration). Since non-compliance is not a random process, an unconfounded measure of effect is difficult to obtain, unless trials are analysed based on the initial randomization, regardless of compliance (intention to treat analysis). This kind of analysis, however, may result in underestimating (or missing) an effect. More importantly, many of the questions we ask in epidemiology do not lend themselves to being studied experimentally, for ethical reasons. We cannot allocate people to an exposure that may cause harm.

It is important that students understand how the three cornerstones of the randomized controlled trial help to isolate the effect of the exposure. Randomization will make the groups comparable at baseline (by reducing confounding) when the study is large. Blinding will make compliance and outcome measurements comparable on average if it can be maintained throughout the trial (no differential misclassification), and placebo use will make circumstances comparable by controlling for the placebo effect. Any differences in the outcome in a well-conducted trial should then be due to the difference in exposure, the difference between the groups to be compared, lack of adherence to the protocol, or chance. Lack of exposure contrast is a major problem if compliance is low or if the drug is available to individuals receiving placebo treatment. More details on the randomized controlled trial are available in the works by Feinstein (1989a, b), Greenland (1990), Senn (1991), Moher et al. (2001), and Friedman et al. (1996).

In cohort studies, we usually want to make the exposure contrast as large as possible to answer the question, does this exposure have an effect or not? In studies in which we wish to estimate the quantitative nature of the association, we usually need to compare several levels of exposure to estimates: at what level does the exposure have an effect, and what shape does the dose–response relation have?

Students often have problems understanding the distinction between effect measure modification and confounding, and it is important for teachers to be aware of this. Effect modification is the phenomenon by which the (theoretical) measure of association between an exposure and a given outcome varies across strata of another determinant (e.g. the incidence of oesophageal cancer is higher than expected (according to whether the model is additive or multiplicative; Rothman 2002) in smokers with high alcohol consumption than in smokers with low alcohol consumption). Effect measure modification is model dependent, and it is always measure dependent. Most measures of association are based on a multiplicative or additive scale, depending upon the model that is chosen to describe the association (rate ratio or rate difference). The presence of a causal effect does lead to effect measure modification on at least one scale.

Effect measure modification, as defined by a deviation from an additive model, may be a main target for study, as proposed by Rothman (2002). This is best studied in a well-balanced design that uses the principles of the factorial design (Armitage and Berry 1994).

Confounding is a type of bias that depends on an *unbalanced* distribution of determinant(s) of the outcome that also cause(s) the exposure directly or indirectly. Thus, the exposed and unexposed groups would have a different disease occurrence even in the absence of exposure. A potential effect modifier may or may not be a confounder, depending upon its association with the determinant under study. On the other hand, an effect modifier is part of the mechanism of the hypothesis under study. DAGs (Pearl 2010), while potentially useful for assessing confounding, are not able to represent effect modification. Furthermore, they do not permit representation of relative strength of effect or dose-response.

We usually use fixed cohorts to study the effect of a given exposure localized in time, calendar time, or age. We use cohorts with an open entry in, for example, studies of occupationally exposed people who enter the cohort when they become exposed. If exposure duration is unrelated to the occurrence of the disease under study and has a short incubation time, follow-up may end as soon as the exposure ends; however, usually members of a cohort are followed up longer, regardless of the duration of exposure. Measurement of the exposure should be updated if the follow-up is long, as is, for example, done in the Nurses'

Health Study. There are some instances in which only a specific window of exposure is investigated, such as exposures during a specific stage of foetal life, childhood, or puberty. In this case, the effort to collect exposures has to be focused on these particular relevant periods. Proper effect measures are rates that take follow-up time from exposure into consideration.

Students at the graduate level should also be taught more complicated designs, such as case-only studies, twin studies, or other family studies.

When introducing the case-control study, it is important to stress that this design tries to capture the information in the underlying population by using a different sampling strategy from that of the classical follow-up study. The aim is to estimate the relative effect measures during follow-up without studying the entire population. Case-control studies are usually more efficient—and cheaper—than cohort studies because they obtain almost the same amount of information with much fewer observations (all cases plus a variable number of controls). We can illustrate the link between a cohort study and a case-control study as follows.

If we imagine an underlying fixed cohort without loss to follow-up, we have the data for all the relative effect measures, as illustrated in Table 3.1.

If we assume that a registry captures all the cases from this underlying population and we are able to reconstruct valid data on the exposures, we can estimate the numerator exposure odds indicated on the right side of these formulas. To obtain the denominator exposure odds, we sample from the cohort at baseline (to estimate relative risk), from the observation time during follow-up (to estimate the incidence rate ratio), or from those who remain disease free throughout follow-up time (to estimate the disease odds ratio).

Table 3.1 Relative effect measures for cohort analysis*

Measure	Formula	Alternative formula
Relative risk	$\frac{a/N+}{c/N-}$	Section 1.01 $\frac{a/c}{N+/N-}$
Incidence rate ratio	(a) $\frac{a/t+}{c/t-}$	(i) $\frac{a/c}{t+/t-}$
Odds ratio	1) $\frac{a/b}{c/d}$	a) $\frac{a/c}{b/d}$

* a = number of diseased individuals in the exposed cohort; b = number of nondiseased individuals in the exposed cohort; c = number of diseased individuals in the unexposed cohort; d = number of nondiseased individuals in the unexposed cohort; N+ = start of follow-up of exposed cohort; N– = start of follow-up of unexposed cohort; t+ = observation time for the exposed cohort; t– = observation time for the unexposed cohort.

A random sample taken at the start of the follow-up provides estimates of the denominators which, for the relative risk, are the exposure distributions in the entire study base at the start of follow-up (N_+/N_-). If the follow-up is based on members of a registry or a biological bank, and the follow-up time is short, a random sample from this study base will estimate the exposure distribution of interest. We call this design a case-cohort study. Its merits are mainly that one control sample may be used for several case groups (Greenland 1986; Prentice 1986). If cohort members are lost over the follow-up time, the analyses have to take this into consideration.

If there is loss to follow-up, or if the study base is a dynamic population, we need to estimate the distribution of exposed and unexposed time during follow-up. If the underlying population is a closed cohort, comprising, for example, members of a biological bank (a repository of biological specimens taken from a given population in a given time period), we could let the members of the cohort be represented by the units of their exposed follow-up time or unexposed follow-up time, respectively, and then sample from among these units. In other situations we would sample from the population at risk at the time when cases were detected and accept that controls maybe sampled more than once and thus also enter the case group if they become cases. We call this design a case-base study or a case-control study with incidence density sampling.

In the case-cohort study we would sample controls that may also include cases, since cases have the same probability of being selected as non-cases. In the second situation we would sample controls that may later become cases. It is thus important to accept that the same subjects may be represented as both cases and controls. Sometimes this requires refinement of the normal statistical procedures but these simple sampling procedures bypass the 'rare disease assumption' often mentioned in textbooks (i.e. that odds ratios only estimate the relative risk in case of 'rare' diseases—that have to be rare within each category of the exposure). In this situation we obtain unbiased estimates of relative risk and incidence rate ratio, regardless of the frequency of the disease. (Some would call a case-cohort study a nested case-control study but, in principle, all case-control studies are nested within a cohort, although the underlying cohort may be difficult to identify.)

Only for the case–non-case study, in which controls are sampled from those who remain disease free during follow-up does the rare disease assumption apply, and only if we want the odds ratio to estimate the relative risk or incidence rate ratio. In this example the exposure odds ratio is the same as the disease odds ratio (i.e. $\left(\frac{a/c}{b/d}\right)$).

A number of designs are based upon case-only studies. The most important of these is the case-crossover study, which is especially useful when trying to avoid confounding by stable time factors. This design is best suited to study immediate treatment effects or effects that happen shortly after the exposure (MacLure 1991; Mittleman et al. 1995; Greenland 1996). If patients use a given drug only occasionally, for example, the incidence rate of a specific short-term side effect of using the drug can be compared with the incidence rate in periods of no treatment. The ratio between these two incidence rates reflects the association between the treatment and the potential side effect, and this information may only be captured by studying the past exposure history of those patients who exhibited side effects. By comparing drug use prior to the outcome with drug use at a reference time period for the same persons, the underlying incidence rate ratios can be ascertained for members of the study. Another important design is the case-parent design, which permits the assessment of genetic effects as well as gene-environment interactions (Wilcox et al. 1998).

Almost all designs that address etiologic problems (i.e. the occurrence of an illness as a function of a given determinant) must be based upon a longitudinal recording of determinants and outcomes, perhaps with the exception of exposures that do not change over time, such as genetic factors or exposures occurring during a specific time window (such as prenatal exposures, or exposures occurring in early childhood or puberty).

Cross-sectional surveys are usually classified in the context of descriptive studies that aim at estimating the prevalence of a specific illness at a given time in well-defined populations. For this type of study, the principles of random sampling from a target population apply. The study aims at making inferences for a specific population, and it is particularistic in the sense that results are valid for that population only and at a given point in time (the findings refer to prevalence, not risk). When we estimate effect measures in etiologic designs, we usually do not have a target population in mind and we expect (or hope) the findings will apply to other populations or for other time periods. If not, causal pathways in different populations include other factors that modify the effect.

Teaching method and format

In our experience, design issues are best taught by using a mixture of different teaching methods. Student participation is of key importance. The concepts and principles needs to be discussed and used in practical examples.

The logical structure of the design, criteria for protocol writing, and good epidemiological practice should be presented in lectures, and the students should read the relevant textbook chapters.

Published articles should be presented to the students and discussed with respect to the design and the authors' conclusions in light of the stated objective. Students should take part in open discussions about the appropriateness of the authors' design, the validity of the study, and the validity of the conclusions. This is usually a successful exercise, as it focuses the students' attention on important aspects of the subject while, at the same time, stimulating critical assessment. It is also preparatory for the following step.

Published papers should be read and discussed by students according to standard criteria for peer review (Elwood 1998; Savitz 2003) and this discussion should be brought forward to a plenary debate.

Students at a more advanced level should be asked to write a protocol on a specific topic and have it discussed in plenary.

It is important at all levels to illustrate the pitfalls in all designs, especially in designs that deviate from routine principles.

Assessing students' achievements

Most students find design problems interesting and take an active part in discussions, often taking a very critical point of view. It should be stressed that a critical approach has to take into consideration what is important and what is less important. There is a difference between being able to discuss the problems and understanding the basic concepts. Since the teaching has to move from basic principles to more advanced topics, it is important to implement an ongoing assessment procedure and make sure that students master a topic before the lectures move forward to the next. At the end of the course, students should be evaluated on the basis of their understanding of the concepts, rather than on the learning of specific details. Oral examinations or written essays may be better than multiple-choice exams at capturing the overall level of understanding of the students. For undergraduate students, we should establish that they have a firm grasp of the various design options and are able to discuss the problems of confounding and bias inherent to these designs. At a more advanced level, students should be able to review study protocols and critically appraise published papers (Table 3.2).

Conclusion

Even students in basic courses in epidemiology should be familiar with the most common study designs. They will not be able to read most epidemiologic papers critically unless they are familiar with at least some of these designs. Conducting epidemiologic research requires not only a more profound understanding of these designs but also experience in what can go wrong in the data collection process.

Table 3.2 Syllabus for ten one-hour lectures followed by one-hour group exercises (twenty hours)

Lecture	Topic
1	Causation, including the use of directed acyclic graphs
2	Measures of disease occurrence and association
3	Design options and design classification
4	The randomized trial
5	Cohort studies
6	Case-control studies
7	Bias and confounding
8	Cross-sectional studies
9	Case-only studies including case-time-control studies
10	Recap of strength and weaknesses in different designs

Notes

Lecture 1. The concept of causation plays a key role in understanding our measures of association and concept of interaction. The lecture should also introduce DAGs (Pearl 2010).

Lecture 2. In this lecture the concepts of risks, rates, proportions, and odds, as well as risk/rate and odds ratios, are revisited.

Lecture 3. We can present the different design options based on a given research question, such as whether hormone-replacement therapy causes breast cancer.

Lecture 4. Although the randomized trial is not a model for the other designs, it is conceptually easy to understand for most students and illustrates important principles that have to be taken into account in all studies.

Lecture 5. The randomized controlled trial is itself a cohort study but students need to be introduced to other cohort studies that do not randomize the exposure. The Framingham study, the Nurses' Health Study, and the Danish National Birth Cohort can all be used as examples.

Lectures 6. Most of the new insight into design methodology addresses the case-control study. This lecture should focus upon their similarities to the cohort study.

Lecture 7. Most method problems can be categorized as confounding selection bias or information bias. Directed acyclic graphs may be used to illustrate the bias.

Lecture 8. Cross-sectional studies/surveys are frequently used, and their strength and limitations in descriptive as well as analytical epidemiology need to be identified.

Lecture 9. New design options, like the case-crossover study and perhaps also the case-parent design, should be presented.

Lecture 10. Again, use examples to illustrate the strengths and weaknesses of the various designs.

References

Ahrens, W. and Pigeot, I., eds. (2013) *Handbook of Epidemiology* (2nd edn). Berlin: Springer.

Armitage, P. and Berry, G. (1994) *Statistical Methods in Medical Research*. Cambridge: Blackwell Science Ltd.

Axelson, O. (1979) The case-referent (case-control) study in occupational health epidemiology. *Scandinavian Journal of Work, Environment and Health*, **5**: 91–9.

Basso, O. (2007) Options and limitations in studies of successive pregnancy outcomes: an overview. *Paediatric and Perinatal Epidemiology*, **21**: 8–12.

Elwood, J. M. (1998) *Critical Appraisals of Epidemiological Studies and Clinical Trials*. Oxford: Oxford University Press.

Feinstein, A. R. (1973) Clinical biostatistics. XX. The epidemiologic trohoc, the ablative risk ratio, and "retrospective" research. *Clinical Pharmacology and Therapeutics*, **14**: 291–307.

Feinstein, A. R. (1989a) Epidemiologic analysis of causation: the unlearned scientific lessons of randomized trials. *Journal of Clinical Epidemiology*, **42**: 481–9.

Feinstein, A. R. (1989b) Unlearned lessons from clinical trials: a duality of outlooks. *Journal of Clinical Epidemiology*, **42**: 497–8.

Friedman, L. M., Furberg, C. D., and DeMets, D. L. (1996) *Fundamentals of Clinical Trials* (3rd edn). St Louis, MO: Mosby, Springer.

Greenland, S. (1986) Adjustment of risk ratios in case-base studies. *Statistics in Medicine*, **5**: 579–84.

Greenland, S. (1990) Randomization, statistics, and causal inference. *Epidemiology*, **1**: 421–9.

Greenland, S. (1996) Confounding and exposure trends in case-crossover and case-time control designs. *Epidemiology*, **7**: 231–9.

Greenland, S., Pearl, J., and Robins, J. M. (1999) Causal diagrams for epidemiologic research. *Epidemiology*, **10**: 37–48.

Grodstein, F., Manson, J. E., Colditz, G. A., Willett, W. C., Speizer, F. E., and Stampfer, M. J. (2000) A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Annals of Internal Medicine*, **133**: 933–41.

Grodstein, F., Stampfer, M. J., Manson, J. E., Colditz, G. A., Willett, W. C., Rosner, B., Speizer, F. E., and Hennekens, C. H. (1996) Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *New England Journal of Medicine*, **335**: 453–61.

Hardin J. W. and Hilbe J. M. (2003) *Generalized Estimating Equations*. Boca Raton, FL: Chapman & Hall/CRC.

Hernán M. A., Alonso, A., Logan, R., Grodstein, F., Michels, K. B., Stampfer, M. J., Willett, W. C., Manson, J. E., and Robins, J. M. (2008) Observational studies analyzed like randomized experiments. *Epidemiology*, **19**: 766–79.

Hernán M. A., Hernández-Díaz, S., Werler, M. M., and Mitchell, A. A. (2002) Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *American Journal of Epidemiology*, **155**: 176–84.

Hill, A. B. (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **58**: 295–300.

Holick, M. F. (2007) Vitamin D deficiency. *New England Journal of Medicine*, **357**: 266–81.

Howards, P. P., Schisterman, E. F., Poole, C., Kaufman, J. S., and Weinberg, C.R. (2012) "Toward a clearer definition of confounding" revisited with directed acyclic graphs. *American Journal of Epidemiology*, **176**: 506–11.

Howards, P. P., Schisterman, E. F, and Heagerty, P. J. (2007) Potential confounding by exposure history and prior outcomes: an example from perinatal epidemiology. *Epidemiology*, **18**: 544–51.

International Epidemiological Association (IEA)/The European Epidemiology Group. (2007) *Good Epidemiological Practice: Proper Conduct in Epidemiologic Research*. <http://www.ieaweb.org.>, accessed 30 October 2014.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., and Davey Smith, G. (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, **27**: 2974–6.

Lewis, D. (1973) Causation. *Journal of Philosophy*, **70**: 556–67.

Mackie, J. L. (1975) *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.

MacLure, M. (1991) The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, **133**: 144–53.

Miettinen, O. S. (1981) Confounding: essence and detection. *American Journal of Epidemiology*, **114**: 593–603.

Miettinen, O. S. (1985a) *Theoretical Epidemiology*. New York: John Wiley & Sons.

Miettinen, O. S. (1985b) Design options in epidemiologic research: an update. *Scandinavian Journal Work, Environment and Health*, **8**: 7–14.

Miettinen, O. S. (1988) Striving to deconfound the fundamentals of epidemiologic study design. *Journal of Clinical Epidemiology*, **41**: 709–13.

Miettinen, O. S. (1999) Etiologic research: needed revisions of concepts and principles. *Scandinavian Journal of Work, Environment and Health*, **25**: 484–90.

Miettinen, O. S. and Karp, I. (2012) *Epidemiological Research: An Introduction*. Heidelberg: Springer.

Mittleman, M. A., MacLure, M., and Robins, J. M. (1995) Control sampling strategies for case-crossover studies: an assessment of relative efficiency. *American Journal of Epidemiology*, **142**: 91–8.

Moher, D., Schulz, K. F., and Altman, D., CONSORT Group (Consolidated Standards of Reporting Trials). (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association*, **285**: 1987–91.

Olsen, J. (1993) Some consequences of adopting a conditional deterministic causal model in epidemiology. *European Journal of Public Health*, **3**: 204–9.

Olsen, J. (2008) Confounding by exposure history and prior outcome. (Letter to the editor), *Epidemiology*, **19**: 635–6.

Olsen, J., Christensen, K., Ekbom, A., and Murray, J. (2010) *An Introduction to Epidemiology for Health Professionals*. New York: Springer.

Olsen, S. F. (1993) Marine n-3 fatty acids ingested in pregnancy as a possible determinant of birth weight: a review of the current epidemiologic evidence. *Epidemiologic Reviews*, **15**: 399–413.

Pearl, J. (2009) *Causality* (2nd edn). Cambridge: Cambridge University Press.

Pearl, J. (2010) An introduction to causal inference. *International Journal of Biostatistics*, **6**: Issue 2, Article 7.

Porta, M. ed., Greenland, S., Last, J. M., associate eds. (2008) *A Dictionary of Epidemiology. A Handbook Sponsored by the IEA* (5th edn). Oxford: Oxford University Press.

Prentice, R. L. (1986) A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, **73**: 1–11.

Rothman, K. J. (1976) Causes. *American Journal of Epidemiology*, **104**: 587–92.

Rothman, K. J. (2002) *Epidemiology: An Introduction* (1st edn). Oxford: Oxford University Press.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008) *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins.

Savitz, D. A. (2003) *Interpreting Epidemiologic Evidence: Strategies for Study Design and Analysis*. New York: Oxford University Press.

Senn, S. J. (1991) Falsificationism and clinical trials. *Statistics in Medicine*, **10**: 1679–92.

Susser, M. (1991) What is a cause and how do we know one? A grammar for pragmatic epidemiology. *American Journal of Epidemiology*, **133**: 635–48.

Szklo, M. and Javier Nieto, F. (2012) *Epidemiology: Beyond the Basics* (3rd edn). Boston, MA: Jones & Bartlett Publishers.

The Women's Health Initiative Study Group. (1998) Design of the women's health initiative clinical trial and observational study. *Controlled Clinical Trials*, **19**: 61–109.

Valeri, L. and VanderWeele, T. J. (2013) Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, **18**: 137–50.

VanderWeele, T. J. and Vansteelandt, S. (2009) Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, **2**: 457–68.

Varas-Lorenzo C, Garcia-Rodriguez LA, Perez-Gutthann S, et al. (2000) Hormone replacement therapy and incidence of acute myocardial infarction. *Circulation*, **101**: 2572–8.

Weinberg, C. R. (1993) Toward a clearer definition of confounding. *American Journal of Epidemiology*, **137**: 1–6.

Weinberg, C. R. (2003) Studying parents and grandparents to assess genetic contributions to early-onset disease. *American Journal of Human Genetics*, **72**: 438–47.

Weinberg, C. R., Shore, D. L., Umbach, D. M., and Sandler, D. P. (2007) Using risk-based sampling to enrich cohorts for endpoints, genes and exposures. *American Journal of Epidemiology*, **166**: 447–55.

Weinberg, C. R. and Umbach, D. M. (1999) Using pooled exposure assessment to improve efficiency in case-control studies. *Biometrics*, **55**: 718–26.

Weinberg, C. R. and Umbach, D. M. (2000) Choosing a retrospective design to assess joint genetic and environmental contributions to risk. *American Journal of Epidemiology*, **152**: 197–203.

Weinberg, C. R. and Umbach, D. M. (2005) A hybrid design for studying genetic influences on risk of diseases with onset in early life. *American Journal of Human Genetics*, **77**: 627–36.

Wilcox, A. J., Weinberg, C. R., and Basso, O. (2011) On the pitfalls of adjusting for gestational age. *American Journal of Epidemiology*, **174**: 1062–8.

Wilcox, A.J., Weinberg, C.R., and Lie, R.T. (1998) Distinguishing the effects of maternal and offspring genes through studies of "case-parent triads." *American Journal of Epidemiology*, **148**: 893–901.

Chapter 4

Statistics in epidemiology

Per Kragh Andersen

Introduction to statistics in epidemiology

Epidemiology deals with the assessment of effects of risk factors on disease outcome in human populations, and statistics has played a prominent role in its development. In fact, some authors (Krickeberg 1992) have claimed that epidemiology is really a part of statistics!

This chapter provides a guide for the 'inexperienced' teacher to teach statistics as a part of a teaching programme in epidemiology. By 'inexperienced', I mean that the reader has little or no experience in teaching statistics in an epidemiologic context and not that they know little about statistics. In fact, I generally recommend that statistics be taught by professional statisticians and not by those from an applied field with a more narrow background in statistics. It is, however, crucial that the potential teacher is familiar with the field of epidemiology to motivate the students and to illustrate points with relevant examples. As will be apparent from the following, I think that teaching any statistical topic for students in epidemiology (and from any other field for that matter) should take relevant motivating examples as the starting point. Nevertheless, I would like to emphasize that, in my opinion, statistics is not well suited to problem-based learning. The reason is that the statistical concepts and methods should to a large extent follow a strict logical order which is difficult, if not impossible, to achieve using problem-based learning. Thus, the course to be described follows a strict deductive model.

I will assume that a course introducing the standard epidemiologic designs is taught concurrently with the statistics course to be described. It does not make much sense to give first a course describing the designs and not the analysis and then subsequently to give the statistics course. Furthermore, the course to be outlined is intended for undergraduates with no prior familiarity with statistics.

Teaching objectives

Since I will assume that this course provides the first teaching in statistics to which the students are exposed, the first main objective is to make them

- understand variation.

Without variation (interindividual, intraindividual, and sampling variation), there would be no need for statistics whatsoever! Next, the students should

- understand the concepts of population versus sample, and parameter versus estimate; which naturally leads on to how to
- understand standard deviations and confidence intervals by discussing repetition of 'experiments'. (Here, I do not think that it is necessary to spend much effort in distinguishing between standard deviation and standard error—it is just a question of which experiment to repeat).

After these basics, the students should learn how to

- estimate probabilities ('risks' and 'prevalences') with confidence limits;
- estimate relative risks (and risk differences) with confidence limits; and
- estimate odds and odds ratios with confidence limits from cohort studies with common follow-up times for everyone and from cross-sectional studies.

Furthermore, the student should learn how to

- compare two probabilities or odds (among 'exposed' and 'unexposed') using the chi-square test;
- understand the concepts of significance level and *P*-value; and
- understand the close connection between significance tests and confidence limits (in these simple two-sample situations).

Next, the students should

- understand rates and intensities as useful measures of disease frequency in cohort studies where everyone does not necessarily have the same follow-up time;
- understand the connection between rates and risks;
- learn how to estimate rates and rate ratios with confidence limits; and
- learn how to compare two rates (among 'exposed' and 'unexposed') using a chi-square type test.

The last introductory topic to discuss is case-control studies, which are conceptually somewhat more complicated than cohort studies or cross-sectional studies (involving as they do a sampling of the population generating the cases). The students should understand

- how a case-control study may be thought of as taking place by sampling within a given cohort (though it is perhaps more natural to discuss this as a part of the concurrent epidemiology course) and how the controls may be selected in various ways. For the sake of simplicity of analysis in a basic

course such as the one described in this chapter, I would recommend concentrating on the kind of 'case–non-case' design which allows estimation of the odds ratio. (This is also the design for which more advanced regression analyses are most simple.)

The student should also

- understand how to analyse the odds ratio based on the same basic two-by-two table as in a cohort study.

At this stage, the students will be ready to learn how to

- understand and estimate a population-attributable risk and see its relevance in a public health context. Exactly when in the later part of the course this topic is taught is probably of minor importance.

The rest of the objectives (and for the rest of the course contents, see 'Teaching contents and format') deal with how to

- understand and adjust for confounding.

This is a large and complicated task and the teacher's ambition may here be on a number of different levels, ranging from a fairly brief discussion of the topic via stratified ('Mantel–Haenszel') analyses to a thorough treatment of, for example, logistic regression. I think that a suitable starting point for the discussion of confounding is that of 'a fair comparison between the exposed and the unexposed group'. That is, if we randomly select exposed and unexposed individuals, does this then provide a fair comparison? Or do we run the risk that perhaps the exposed individuals are older or smoke more than the unexposed individuals? If the latter is the case, then some sort of adjustment for age and smoking is required (provided that age and smoking are associated with the disease risk). This way of thinking is very much in line with the discussion of (randomized) experiments given by Clayton and Hills (1993) and leads very naturally to the concept of a stratified analysis. A stratified analysis may be performed for all the measures for the comparison mentioned above (i.e. relative risks, odds ratios, rate ratios, etc.), and the idea is that the comparison is made between randomly selected exposed and unexposed individuals from the same stratum (e.g. age group) and a common (Mantel–Haenszel) estimate is then obtained by summarizing estimates from the individual strata. The student should be able to

- understand the rationale behind stratified analyses, and to
- perform a stratified analysis for each of these measures.

This includes

- computing the Mantel–Haenszel estimator with confidence limits and the corresponding test statistic and

- examining whether the modelling assumptions behind the analysis are reasonably fulfilled, that is, whether the measure is approximately constant over strata).

This leads to the requirement that the students should

- understand the concept of interaction.

However, students should not necessarily be able to understand the details of, for example, the Breslow–Day test for homogeneity of odds ratios across strata; they only need to be able to interpret the result of such a test.
Last, but not least, the students should

- be able to read epidemiological literature which uses and refers to the concepts outlined above.

This really leads to a dilemma, since most modern epidemiologic papers apply some sort of multiple regression analysis such as logistic regression or Poisson or Cox proportional hazards regression, and these techniques are rather too difficult to be treated thoroughly in an introductory course such as the one outlined here (depending, of course, on how much time is available for the teaching). Some compromise is required where the students are introduced to multiple regression analysis and taken through typical tables from epidemiologic papers, hopefully leading to a basic understanding of the fact that regression coefficients and their associated confidence intervals are just like (log) odds ratios which are mutually adjusted as in a series of stratified analyses. A particular problem to deal with in this connection is effects of quantitative explanatory variables that do not have exact analogues in Mantel–Haenszel analyses.

A topic dealt with in most introductory courses on design and analysis of epidemiologic studies is 'standardized rates', which belongs naturally under the heading of adjustment for confounding. Thus, a directly standardized rate, being simply a weighted average of age-specific rates, is computed when age confounding is suspected in a comparison of rates from different groups. Though such a comparison can be performed more directly using stratified analysis, directly standardized rates may be useful, for example, as a descriptive one-number summary of each group. The analogous indirectly standardized rate (and the equivalent standardized mortality ratio (SMR)) is useful when comparing observed rates with standard (population) rates. Again, some sort of age adjustment is necessary and, under the simplifying assumption that the observed age-specific rates are proportional to the population rates (an assumption which may be checked), the SMR is a suitable measure of the common rate ratio. Thus, the student should

- understand the principles and limitations of standardization.

In summary, there are a number of concepts that the student should understand and be familiar with:

- variation, population versus sample, parameter versus estimate, standard deviation/confidence limits, risk, prevalence (probability), odds, rate (intensity), control sampling, population-attributable risk, confounding, and interaction.

There are also a number of techniques that the student should be able to apply:

- estimate risk, prevalence, odds, and rate (and the corresponding ratios for two groups) with confidence limits; test whether ratios are unity using chi-square tests; perform stratified analyses; and compute standardized rates.

In addition, there are some topics for which the student should understand the principles as applied in the literature:

- test for interaction, and multiple regression analysis (including linear effects).

Teaching contents and format

The teaching format will heavily depend on the availability of computers in the classroom. The ideal situation is one in which students have constant access to computers with a statistical package like R, SAS, SPSS, or STATA installed. Here STATA seems to provide a nice compromise between being easily accessible for non-specialists and being sufficiently comprehensive for most purposes. However, the freely available R package is also a strong candidate. This will provide the possibility of mixing lectures with exercises in which students try the methods just discussed on real datasets. Alternatively, lectures may take place in a classroom without computers but the students should then have access to computers elsewhere in order to work with day-to-day exercises illustrating the methods taught during the lectures. The lectures to be described in the following are assumed to be supplemented by

- an introduction to a statistical computer package, and
- a series of computer exercises (in one of the formats just described).

Each lesson should be completed in about three to four 45-minute lectures, plus computer exercises.

Lesson 1: variation and confidence intervals

Introduce this lesson by giving the students a general idea about what statistics is about (i.e. drawing inference on scientific problems from data) and turn quickly to a simple introductory example to ease the understanding of population and

sample. I have sometimes used the following example but, obviously, any similar simple example would do. Imagine that you want to assess the prevalence of asthma in a well-defined geographical region among teenagers, based on a questionnaire. Based on this example, one may discuss population, sample, sampling variation, and interindividual and intraindividual variation. Furthermore, such an example should make the students feel the need for a confidence interval for the true asthma prevalence and should be taught how to do that from the estimated standard deviation of the proportion in the sample. (However, they need not understand why the standard deviation is computed the way it is.) The interpretation of the standard deviation for the distribution of prevalence estimates, and that of a confidence interval through hypothetical repeated sampling, should be explained and, if possible, illustrated on a computer. In a similar vein, risk parameter estimates for exposed and unexposed individuals from a cohort study could be discussed with standard deviations leading to the wish to make a statistical comparison of the two.

Lesson 2: comparing risks and odds

Begin by repeating the highlights of lesson 1, emphasize what the students are supposed to understand and what they have to accept; then go on to introduce the relative risk and an associated confidence interval. Repeat the interpretation of a confidence interval and examine whether the 'null value' RR = 1 is inside or outside the confidence limits in your example. Next, based on that example, introduce the concepts from significance testing: null hypothesis, test statistic, distribution of test statistic, and *P*-value. I would recommend using the Mantel–Haenszel version of the chi-square test in a two-by-two table because, for this statistic, it is quite easily seen how it measures discrepancies from how one would expect the data to look if the null hypothesis were true. Furthermore, it easily generalizes to stratified analysis and to comparison of rates. The students should understand the need for a standard deviation for the difference between observed and expected but they need not understand why the formula looks the way it does. They should understand how a distribution of test statistics under the null may be obtained through repeated sampling and how one then compares the observed value of the test statistic with this distribution (the chi-square distribution with one degree of freedom) and obtains the *P*-value using a table of the distribution. Finally, they should understand the close connection between confidence limits and test statistic (in this simple one-degree-of-freedom situation). The students should also be introduced to the odds ratio (with a confidence interval) and it should be emphasized that the test statistic can just as well be seen as a test for the hypothesis that the odds ratio is unity.

Lesson 3: rates and case-control studies

Begin by repeating the highlights of lesson 2, emphasize what the students are supposed to understand and what they have to accept, and promise them that all the concepts introduced in connection with significance tests will be repeated several times later in the course! Next, introduce the rate as a frequency measure useful in cohort studies with unequal follow-up times and explain how to compute confidence intervals for rates and rate ratios. The connection between a rate (i.e. a constant intensity) and a probability (risk) should be explained. Next, discuss the Mantel–Haenszel type test for comparing two rates and repeat the concepts from significance testing. At this stage, case-control studies can be discussed, emphasizing how odds ratios may be estimated based on observation on cases and a sample of disease-free controls. One could also mention the alternative design where cases are compared to a random sample from the whole population, enabling estimation of the relative risk but complicating analysis due to lack of independence between cases and 'controls' (e.g. Schouten et al. 1993).

Lesson 4: adjustment for confounding using stratified analysis

Begin by repeating the highlights of lesson 3, emphasize what the students are supposed to understand and what they have to accept, and go on discussing confounding. This may, for instance, be done by presenting a (perhaps hypothetical) situation where an exposed and an unexposed group are to be compared and where, for some reason, the exposed group is older than the unexposed. In this case individuals randomly selected from the exposed group and from the unexposed group will not provide a fair comparison, since differences between the risks in the two groups may to some extent be ascribed to effects of age (if age is a risk factor for the disease outcome). Thus, to compare like with like, an age stratification of the two groups is needed, and this leads very naturally to a discussion of stratified (Mantel–Haenszel) analysis of, for example, the odds ratio. The students should understand that the ingredients of a stratified analysis are just what one would have used for comparisons of exposed and unexposed, stratum by stratum (i.e. observed, expected, and standard deviation of observed) but that, instead of performing several analyses, one for each stratum, the three sufficient statistics are added over strata, enabling one to make a single comparison between exposed and unexposed, adjusted for the stratification variable (the confounder). Technically, students should learn about the Mantel–Haenszel estimator, its associated confidence interval, and the Mantel–Haenszel chi-square test. Students should understand

that the estimation procedure corresponds to the calculation of a weighted average of the stratum-specific odds ratios and that it therefore only makes sense if all the stratum-specific odds ratios are 'similar', that is, if there is no interaction. They should know that tests for no interaction exist and how to interpret them, but the details could be left out of this introductory course (depending of course on the available time).

Lesson 5: standardization and attributable risks

Begin by repeating the highlights of lesson 4, emphasize what the students are supposed to understand and what they have to accept, and repeat the basic idea in the stratified analysis by showing how a Mantel–Haenszel analysis of rates is performed. At this stage, standardization techniques can be introduced. The idea behind 'direct' standardization (calculation of a weighted average of age-specific rates) is easily explained but the students should understand that, for a comparison of two groups, an external standard age-distribution is not needed. The principle of 'indirect' standardization should be explained, as well as the way in which this method may be used when comparing rates from the sample with standard ('population') rates. Estimation of the standard deviation of the SMR should be presented. Finally, the population-attributable risk should be introduced, together with both its calculation and its interpretation.

Lessons 1 to 5 as they have now been described provide the necessary minimum content of a basic course in epidemiological statistics but, to fulfil the requirement that the students should be able to read epidemiological articles, some introduction to multiple regression analysis (e.g. logistic regression) should be given. Thus, I would recommend that two extra modules be added with this aim.

Lesson 6: introduction to logistic regression

Begin by repeating the highlights of lesson 5, emphasizing what the students are supposed to understand and what they have to accept. To introduce logistic regression, begin by considering tables. First, in a two-by-two table with one binary outcome variable (diseased versus not diseased) and one binary explanatory variable (exposed versus not exposed), the odds ratio may be calculated as a measure of the effect of the explanatory variable on the outcome. The log odds ratio may also be seen as the coefficient b in the regression model

$$\ln(\text{odds}) = a + bZ$$

when the explanatory variable Z is coded as 0 or 1. Here one may need to remind students of their, perhaps forgotten, knowledge of logarithms from high school.

Next, in a three-by-two table, again with a binary disease outcome and with an explanatory variable with three ordered categories, choosing a reference category enables one to calculate two log odds ratios as measures of the effect. These log odds ratios may be seen as the coefficients b_1 and b_2 in the regression model

$$\ln(\text{odds}) = a + b_1 Z_1 + b_2 Z_2$$

with the explanatory variable coded using indicator variables Z_1 and Z_2 for the two levels of the variable not chosen as reference.

Finally, when the three ordered levels are denoted by $Z = 0$ (reference), by $Z = 1$, and by $Z = 2$, respectively, the coefficient b in the model

$$\ln(\text{odds}) = a + bZ$$

may be seen as a common log odds ratio for 1 versus 0 and for 2 versus 1 (i.e. if the log odds are plotted against Z, a straight line is obtained). This latter model is easily extended to one with a truly quantitative covariate. These are the simplest situations with either one binary, one categorical, or one quantitative explanatory variable, and the tables are seen to be equivalent to simple logistic regression models.

To introduce multiple logistic regression, begin by considering two two-by-two tables where the relation between the exposure Z and the outcome is studied in two strata defined by a third binary explanatory variable X. In this case the Mantel–Haenszel estimator may be calculated as a measure of the effect of Z adjusted for X. Now, rearrange the tables such that the roles of Z and X are interchanged and the Mantel–Haenszel estimator for the effect of X adjusted for Z may be calculated. Finally, we can study the model

$$\ln(\text{odds}) = a + bZ + cX,$$

which includes both explanatory variables. In this model, it can be seen that b is the log odds ratio for Z both when $X = 0$ and when $X = 1$ and that c is the log odds ratio for X both when $Z = 0$ and when $Z = 1$. Thus, this model corresponds to performing both Mantel–Haenszel analyses at one go and thereby mutually adjusting the effects of Z and X. Finally, explain briefly how interaction may be modelled within the logistic regression framework.

With these building blocks, any multiple regression model can be explained by considering one variable at a time and interpreting the corresponding regression coefficients as log odds ratios adjusted for the other variables in the model. Note that by simply adding the terms for the different building blocks, no interaction is tacitly assumed and interaction is only studied when the researcher explicitly introduces it into the regression model.

Lesson 7: literature examples using logistic regression

Begin by repeating the highlights of lesson 6, emphasizing what the students are supposed to understand and what they just have to accept. The purpose of lesson 7 is to discuss a couple of epidemiologic articles that use logistic regression. The first of these should, preferably, be one where the data are documented in fair detail with tables and where classical Mantel–Haenszel techniques are used as an introduction to the multiple regression analyses. If access to the raw data is possible, then reconstructing some of the tables from the article in the computer sessions will usually make an important pedagogic point.

A number of simple statistical methods that really do belong in a basic course in statistical epidemiology have not found their way into these lessons. These include analysis of quantitative outcomes (e.g. *t*-test and linear regression), analysis of survival data (the Kaplan–Meier estimator and the log rank test), analysis of general *r*-by-*c* contingency tables, and analysis of matched data. If time permits, then these are all candidates for inclusion in further lessons.

Assessment of students' achievements

To assess the students' achievements, I would recommend a written exam where some studies and their key results in tabular form are presented to the students, who are then asked to comment and conclude, possibly by doing some computations based on the tables. The latter would be appropriate, especially if computer access at the exam is possible. Otherwise, key results may be given with the tables for comments.

Bibliography

It is not easy to find a textbook with a philosophy exactly as the one outlined in this chapter. However, books which, in principle, cover the material described include older texts like the ones by Sakai and Khurshid (1996) and McNeil (1996), and the more recent book by Woodward (2005). More comprehensive background material for the statistical teacher may be found in Clayton and Hills' (1993) book and in the books by Rothman et al. (2008), Jewell (2004), and Woodward (2005). As mentioned, suitable computer packages to be used by the students include R (available at <http://www.r-project.org>), SAS (available at <http://www.sas.com>), SPSS (available at <http://www-01.ibm.com/software/analytics/spss/>), and STATA (available at <http://www.stata.com>).

Clayton, D. and Hills, M. (1993) *Statistical Models in Epidemiology*. Oxford: Oxford University Press.

Jewell, N. P. (2004) *Statistics for Epidemiology*. London: Chapman & Hall/CRC.

Krickeberg, K. (1992) Moderne Epidemiologie and ihre Anwendungen. *Mitteilungender Deutschen Akademieder Naturforscher Leopoldina*, **35**: 149–60.

McNeil, D. (1996) *Epidemiological Research Methods*. New York: Wiley.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008) *Modern Epidemiology* (3rd edn). Philadelphia, PA: Lippincott-Raven.

Sakai, H. and Khurshid, A. (1996) *Statistics in Epidemiology. Methods, Techniques and Applications*. New York: CRC Press.

Schouten, E. G., Dekker, J. M., Kok, F. J., Le Cessie, S., Van Houwelingen, H. C., Pool, J., and Vanderbroucke, J. P. (1993) Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Statistics in Medicine*, **12**: 1733–45.

Woodward, M. (2005) *Epidemiology: Study Design and Data Analysis* (2nd edn). London: Chapman & Hall/CRC.

Chapter 5

Teaching a first course in epidemiologic principles and methods

Kenneth J. Rothman and Sherri O. Stuver

Introduction to teaching a first course in epidemiologic principles and methods

The first course in epidemiologic methods occasionally serves as the introduction to epidemiology for a student but more typically is preceded by a more general introductory course. Nevertheless, students taking their first course in epidemiologic methods are still green when it comes to first-hand experience with epidemiologic research. More so than courses in language, music appreciation, or mathematics, few students in a first course in epidemiologic methods will have a clear concept of what constitutes the subject matter of the course. Most of them, at best, will have a hazy understanding of the goals of epidemiologic research, and little insight into the principles and methods that they are about to confront. Some will not even have read many, if any, epidemiologic papers; others may have had just enough experience in epidemiology to be dangerous.

Lack of epidemiologic experience poses a barrier to effective learning. People learn more effectively by doing things, under guidance, than by simply listening. Rousing students to action, however, is easier for some topics than for others. When the subject matter is an abstract set of ideas, as is the case for epidemiologic methods, it is difficult to avoid the prototypical lecture scene, a classroom engagement in which the instructor spews abstruse and airy concepts at the audience. This approach can have the same effect as spraying the audience with tear gas: the recipients disperse, enraged at the interaction and hardened in their resolve to cling to their former ideas. It is no easy task, however, to devise an action syllabus to explain concepts such as confounding or non-differential misclassification.

The challenge is to engage the student in activities that breathe life into such abstract concepts. In a chemistry or physics class, laboratory exercises can be

constructed that simulate chemistry or physics research. Unfortunately, conducting or simulating actual research is more difficult for an epidemiology class. Nevertheless, the teacher must try to impart some familiarity with research problems to the students, using laboratories or workshop sessions and exercises to flesh out the abstract ideas.

Some teachers of epidemiologic methods attempt to start students with high-level programming and data analysis even as they flounder in the basic concepts of epidemiologic measurement and study design. Their hope may be that a 'hands-on' approach to the daily activities of computer modelling and other high-tech methods will accomplish the goal of spurring students into action by giving them a glimpse of the research goal and a taste for the steps involved in reaching that goal. However laudable the strategy, teaching students to run software programs is no substitute for an ordered elaboration of the principles and methods of epidemiology. Computers are best left out of the teaching, especially in the early stages. All the computation necessary for teaching epidemiologic methods, at least during the first year or so, can be done handily without any computers at all, beyond a scientific calculator.

In this chapter we suggest a sequence of core topics to broach the teaching of epidemiologic principles and methods. They cover the essential principles and the main methods for conducting epidemiologic research. The topics are causation and causal inference, epidemiologic measures, types of epidemiologic studies, principles of good study design, principles of epidemiologic data analysis, stratified analysis, analysis of interaction, multivariable analysis, and analysis of multilevel or continuous exposures. For each topic, we offer suggestions for what ought to be the focus in the classroom.

Teaching objectives

Students do not become researchers after a single course, nor is the objective of studying the methods of epidemiology necessarily to transform students into researchers. Many courses pose as the objective in such a course the goal that the student can read and comment intelligently on published studies. This is a reasonable goal. The main obstacle in achieving it is that, after a student has acquired a small amount of exposure to epidemiologic methods, there seems to be no study that can stand up to the criticism that will then be heaped upon it. Finding fault seems easy but obtaining the wisdom to know which errors are serious and which are inconsequential is much more difficult to teach. A sound teaching objective would be to have the students able to discuss and criticize the design of epidemiologic studies in a balanced way and to demonstrate quantitative regard for various sources of error.

Teaching content and format

Causation and causal inference

Many notions of causation abound. It is important in a course that will attempt to teach students how to evaluate causal hypotheses to begin with a common understanding of what a cause is. An approach we favour is to introduce the sufficient/component cause model (Rothman 1976) at the beginning of the course and to rely on it repeatedly as the course proceeds. Though it is far from the only useful construct for understanding causation, this model provides a solid conceptual base and can be related to other causal models (Flanders 2006; VanderWeele and Robins 2007). The sufficient/component cause model helps to explain several important epidemiologic concepts:

1. The causes that we study in epidemiology are components of sufficient causes, not sufficient in themselves. That is why smoking is a cause of lung cancer and yet not every smoker gets lung cancer.
2. Component causes interact with one another to form sufficient causes; this interaction is a biologic interaction, involving causal mechanisms (as opposed to statistical interaction, to be discussed later).
3. Factors that interact with one another in a single causal mechanism can do so even if they act at widely differing times.
4. Induction time is the time that represents the accumulation of other, complementary, component causes; induction time characterizes a cause-effect pair, not the disease alone.

And so forth. These conceptual issues can all be illustrated with causal 'pies'.

How can students be drawn into this lesson? One can start by recounting how concepts of causation develop in infants and toddlers, by early experience and genetic programming. Having a theory of causation is a useful survival skill, so it is not surprising that some version of causal thinking develops early in life. Nevertheless, most causal thinking is predicated on the notion of a one-to-one correspondence between cause and effect, a notion that the sufficient/component cause model dispels. Notions of cause are rarely taught in classes before the first course in epidemiologic methods, so students will typically still harbour the concepts that derive from their earliest causal thinking. One can show by example how that thinking evolves during early life, and the ways in which our early concepts of causation tend to be naive. The class succeeds by drawing on insights from child development, psychology, evolution, logic, and philosophy, and by demolishing some cherished misconceptions, a process that is often a good motivator. Exercises can include the fabrication of examples of causal mechanisms to illustrate the various teaching points and can be conducted either individually or in groups.

It is also useful to present the philosophy underlying the scientific method, under the heading of 'causal inference'. The absence of agreement about how causal inference works can be noted, citing sceptics such as Hume and Feyerabend but also highlighting Popper's emphasis on conjecture and refutation, as he stresses the value of formulating hypotheses that are falsifiable (Rothman et al. 2008). Although many have argued against Popper's construction of the scientific method (and in Feyerabend's case against all methods (Feyerabend 1975)), conjecture and refutation brings important value to the scientific process: it encourages the design of studies that illuminate the intersection of competing theories, the point where the theories make differing predictions. Epidemiologic studies that can distinguish between competing theories will accelerate scientific understanding much more rapidly than scattershot approaches that do not tie the studies closely to competing causal mechanisms. Of course, competing theories in epidemiology are often non-causal theories (such as confounding, selection, or information bias) that compete with a causal explanation for an observed finding.

This material may seem dry and can be challenging to teach. In a course on epidemiologic methods, however, a basic introduction to the scientific method is indispensable for a strong orientation. It is most helpful if the principle of choosing among competing theories is re-emphasized along with notions of component causes throughout the course. A useful exercise is to offer examples of competing theories to students and have them propose epidemiologic studies that will refute at least one of the competing theories. One such example concerns the competing theories formulated in the 1970s to explain the reports that use of then new, highly absorbent tampons were a strong risk factor for toxic shock syndrome: one theory suggested that the tampons acted as a culture medium for pathogenic bacteria, which produced the toxin, while a competing theory proposed that the toxin was a chemical manufactured in the new tampons. The key epidemiologic observation is the trend in risk of toxic shock syndrome according to the length of time a tampon was used before a new tampon was inserted. The two competing theories would predict opposite trends, so an epidemiologic study that examined this trend would be able to refute at least one of these theories.

As stated above, it is useful to challenge cherished misconceptions head on. A key misconception that should be addressed early in the course is that epidemiologic studies should enrol study subjects who are representative of broader target populations (Rothman et al. 2013). Many students arrive with the belief that representativeness is essential for generalization. This notion can be challenged in several ways. First, one can emphasize that the concept of statistical generalization, from sample to population, is different from that of

scientific generalization, which is better described as the process of formulating and testing hypotheses about nature. Second, one can contrast the conduct of survey research to the conduct of laboratory studies on animals. Surveys, which are conducted to take polls, obtain marketing information, or even assess health, do depend on representativeness for inference but their findings are not scientific inferences. Rather, the findings are specific to a single time and place. In contrast, laboratory studies on animals such as hamsters, mice, and rats provide scientific inferences that may apply to human health and disease, despite the fact that the hamsters, mice, and rats in the studies are not even representative of all hamsters, mice, and rats, much less the humans for which the inferences are targeted. Some students will let go of ingrained misconceptions only reluctantly, but, successful or not, spotlighting these misconceptions early in the course will ignite interest in the course material.

Epidemiologic measures

Having dealt with the philosophic building blocks, one can turn to epidemiology itself, beginning with epidemiologic measures. The basic presentation comprises measures of disease frequency and measures of effect. For disease frequency, the crucial distinction is between risk and rate (Morgenstern et al. 1980). The way in which time is handled for both these measures must be understood by students. There should also be a clear idea of the meaning of competing risks. Again, these concepts may seem dry. The teacher needs to liven it up with examples and by engaging students in problems that involve each of the measures. Have students compute the risk of dying for members of a small, fixed population. Show them that the risk is near zero over a short period, climbs with time, and how it reaches 100 per cent given enough time. Giving students an exercise in which they calculate rates with a hand tabulation of person-time allows them to go through the task of classifying person-time and events for a set of individuals. Doing so will crystallize the process that underlies an essential epidemiologic concept. As for competing risks, students should be able to explain why a battlefield is not the right spot to conduct a campaign against smoking, and they should be able to explain why risk of death from specific diseases is not a directly observable measure.

The teaching of epidemiologic measures may evoke dreary memories of the fuzzy distinctions from physics classes between force and acceleration or other physical measures that stymied students earlier in their education. Epidemiology teachers face an even greater challenge, because epidemiologic measurements themselves are hard to come by; it is usually easier to measure a force or a mass than to measure the rate of disease. Presenting students with raw data on

individuals and asking them to derive measures from those data will reduce the level of abstraction. Giving them exercises in which the conclusions depend entirely on a correct understanding of epidemiologic measures will engage them even further (Crombie and Tomenson 1981; Hulka et al. 1981; Horwitz et al. 1981; Merletti and Cole 1981).

Types of epidemiologic studies

There are many taxonomies for epidemiologic studies, and classifying some newer study types may add confusion. It is best to opt for simplicity over complexity, emphasizing the main dichotomy, cohort studies versus case-control studies. The prototype for a cohort study is the randomized controlled trial, which is valuable to present both as a paradigm and because it is an important and popular type of study. Nevertheless, one can argue that the conceptual understanding of the case-control study is a greater achievement for epidemiologic methods. Traditionally, case-control studies have been taught as a type of logically inverted cohort study, where the scientist, rather than looking forward from cause to effect, looks backward from effect to cause. A more enlightened view is that a case-control study is a cohort study with an efficiency gain that comes from taking a sample of the denominator experience rather than having to observe the entire denominator experience. The connection between cohort studies and case-control studies can be illuminated by characterizing every case-control study as being nested within a cohort. The definition of the cohort is implied by the definition of the cases in the case-control study; ideally, the controls are sampled from the same people who could have become cases in the study and during the same time that they could have developed the disease. In practice, controls may be sampled from outside this population but the study can, nevertheless, be valid as long as the proxy population gives the same exposure distribution as the actual source population for cases. By emphasizing such parallels between a cohort study and a case-control study, the teacher can provide a conceptual basis for one of the most perplexing design issues, namely, the suitability of a given type of individual as a control in a case-control study.

If cohort studies always measured risks rather than rates, then the corresponding case-control study would be easy enough to comprehend, except for the stumbling block that controls, being sampled from the denominators of the risk measures in a cohort study, would include some individuals who developed disease. The teaching trick here is to make clear that the sampling can be understood as coming from the cohort at the start of its follow-up, when everyone in the cohort was free of disease. Thus, at the time of the sampling, the sampled controls were all free of disease, despite the fact that some of them

may have gone on to become cases. It is essential when presenting these concepts to emphasize that the cases in a case-control study correspond to the numerator of a risk or rate and that the controls correspond to the denominator. In every cohort in which risk is estimated, any case in the numerator of the risk proportion is also represented in the denominator; by analogy, the corresponding case-control study should not aim to exclude future cases from a control series.

Cohort studies often measure rates, however, rather than risks. Then the parallel with case-control studies becomes more difficult to explain, because each control now represents not a sample from a number of people but a sample from an amount of person-time. The teaching issue is the concept of density-based sampling in case-control studies, and the related topic of risk-set sampling. A useful teaching example to think through is a case-control study of the relation of bicycle helmets to head injuries among cyclists (Thompson et al. 1989). The students can be asked to consider what cohort study they would design to evaluate the problem. The question devolves to a consideration of the time at risk, which for helmet-wearing is clearly only the time spent wearing the helmet while riding the bicycle. (Some would restrict the study base even further, to include only the time of an accident, on the theory that a helmet could only prevent an injury during an accident. This restriction, however, ignores the possible effect of a helmet on the risk of an accident, such as by restricting visibility while riding or by emboldening the rider.) In a cohort study one would ideally calculate rates based on time spent cycling with or without helmets. The class can then focus on how to design a case-control study in which controls can be sampled to give an estimate of the distribution of time spent by riders on a bicycle with and without a helmet. The range of possible designs and the inferences that they allow offer considerable insight into the conceptual underpinnings of the case-control study.

Principles of good study design

An indispensable theme throughout the course is the theme of measurement. It is lamentable that much of data analysis is devoted to statistical significance testing, which is not measurement at all but flawed decision analysis. To elaborate the principles of study design, one starts from the premise that the aim of the research is to obtain a valid and precise estimate of an epidemiologic measure. The principles of good study design then become synonymous with obtaining an accurate measure. In measurement, and in epidemiologic studies, one aims to reduce both random and systematic error.

Too much emphasis is put on random error, primarily through significance testing in data analysis but also in planning study sizes by performing power

calculations. The usefulness of power calculations is overrated. The most important message is often obscured in discussions of power or study size, which typically imply that studies are either large enough or not large enough. A better concept is simply that larger studies provide more precise measurements than smaller ones, along a continuous curve. Some study designs are more efficient than others but size is usually the main determinant of the amount of information that a study will provide.

The core discussion with respect to the principles of good study design relates to systematic error, or bias. Most texts take the approach of dividing biases into three broad types: selection bias, information bias, and confounding. These are useful distinctions to maintain. Students often exhibit a reasonable intuitive feel for selection bias. Confounding, too, while having its measure of subtlety, can also be readily digested by most students.

A useful example to introduce confounding involves an illustration of a strong effect of birth order on Down's syndrome prevalence, and then a stronger effect for mother's age that accounts entirely for the birth order effect. One then can emphasize that, because mother's age is just a marker for the occurrence of other biological processes that occur while the mother is ageing, it must therefore also be confounded by as yet unidentified factors. This approach presents the research process as a gradual peeling-off of layers, as each apparent effect is explained by another. The lesson is that, while confounding is considered a bias (and therefore something to be avoided or removed), it is also commonplace and to some extent inevitable. Nearly every causal mechanism raises possibilities for confounding that ought to be considered. It may also be useful to present an example of Simpson's paradox, in which confounding is strong enough to reverse the direction of an association (Reintjes et al. 2000; Hernán et al. 2011).

The effects of non-differential misclassification are often the least intuitive of the principles in this section of the course. It is important to emphasize that bias resulting from non-differential misclassification is, if anything, more ubiquitous than confounding, and, like gravity, is always tugging in the same direction, at least for dichotomous exposures. Even inexperienced students can readily name myriad sources of non-differential misclassification. An important one that they might not offer is mis-specification of the induction time in the study hypothesis. Many students might believe that, since most studies do not bother to specify any induction time hypothesis, the study cannot have misspecified the induction time. The teaching goal here is to lead students to the point where they appreciate that every analysis implicitly rests on some induction time hypothesis, even if it is unstated. To the extent that the hypothesis is wrong, for example, asking about ever-use of aspirin when aspirin affects

disease risk for only twenty-four hours after ingestion, the results can be extremely biased. It is equally important as a teaching goal to help students appreciate that even serious non-differential misclassification cannot explain a strong non-null finding (again, assuming a dichotomous exposure). A good teaching example is the criticism of studies showing a strong relation between spermicides and birth defects; a major criticism was that the underlying data on spermicide use, based on prescriptions rather than on actual use, were a poor proxy for exposure. A clear understanding of the principles of study design should lead the students to understand not only why the criticism is wrong, but why it would have been important if the study had found no effect.

Principles of epidemiologic data analysis

A principle that we emphasize strongly in our courses is the preferred use of estimation rather than significance testing. Examples of mistakes stemming from mindless applications of statistical significance testing are legion, and are useful. Although many teachers seem to urge students simply to be 'more careful' in their use of significance testing, we teach that there is no use for statistical significance testing in epidemiologic analysis. It is perfectly possible to write excellent papers, which will be acceptable to the best journals, without a single test of significance or even a single *P*-value. Estimation methods not only convey more information than *P*-values or significance testing, but they also help avoid the pitfalls that come with testing.

Despite emphasis, the key conceptual point is sometimes lost on students. Many defenders of significance testing point out that a confidence interval and a *P*-value have mathematical links, implying some sort of equivalence. True enough: one can use confidence intervals as surrogate tests and thereby lose nearly all the value in having obtained the estimate. The real distinction is in how the estimate is interpreted, which should be quantitatively rather than qualitatively. If a confidence interval is used as a test to assess whether the null value lies within the interval, the interpretation is qualitative. Instead, one can present the confidence interval function as a tool for visualizing a quantitative interpretation for the estimate. With a quantitative interpretation, the exact location of the boundary of the interval or the exact level of confidence does not matter. It may seem less precise to avoid coming to a qualitative conclusion but it is more quantitative than testing because it allows gradations of interpretation rather than all or none (Poole 1987; Lang et al. 1998).

From this point on in the course, as analytic topics are explored more deeply, quantitative interpretations should continue to be emphasized. Typically, it is only with the greatest reluctance that students, especially the more experienced students, overcome their reliance on statistical significance testing. There appears

to be three components to the enduring fondness with which students embrace statistical significance testing:

1. Before one learns anything at all about statistics or science, one hears about results that are 'statistically significant'. The phrase has such an authoritative ring that it must encourage students to covet work that would likewise be 'statistically significant'; it has marketing appeal.
2. Since interpreting data can be hard work, it would seem much easier to rely on labels such as 'statistically significant' or 'not significant'. The reason, of course, is because one is no longer going through the effort to interpret the data, so, of course, it is much easier.
3. Finally, everywhere one turns, including statistics courses, other epidemiology courses, and most published work in the health field, students will see a conventional reliance on statistical significance testing being reinforced.

To swim against the tide with regard to significance testing is not easy and few teachers attempt it. Nevertheless, for those teachers that succeed in delivering this message, it could be the most important advantage that they bestow on the students on their courses. Perhaps the biggest didactic obstacle is the third point in the above list. Many students object to dissonance between courses. It is helpful to stress that there is little reason for different courses to be harmonious when the world at large is not: students must learn to evaluate divergent views on their merits. Students may be shy about challenging the teaching despite lingering objections relating to the list above. To stimulate discussion, it helps to bring up relevant arguments and address them without waiting for the questions to be raised by students.

Stratified analysis

After elaborating the basics of calculating point and interval estimates, the core of epidemiologic analytic thinking can be conveyed in a thorough discussion of stratified analysis. It is at this time that one can introduce the concept of 'effect-measure modification'. Stratified analysis is the primary tool (or more accurately, should be the primary tool) for evaluating and controlling confounding, as well as for describing effect-measure modification. These two phenomena are occasionally confused by students, who often ask whether a third variable can be both a confounder and an effect-measure modifier at the same time.

Our approach to stratified analysis is to keep the evaluation and control of confounding as separate as possible from the description of effect-measure modification. To begin with, emphasize that, when confounding occurs, it is the crude effect estimate, obtained without stratification, that is confounded. Stratified analysis to control confounding aims to substitute that crude estimate with

an unconfounded replacement. The Mantel–Haenszel version of a pooled estimate is simple and statistically well behaved and should suffice in a first course on methods as the only pooled estimate to consider. The Mantel–Haenszel estimate becomes a replacement for the crude estimate, and the difference between (or ratio of) the two estimates is a measure of the amount of confounding. A discussion of standardization, which does not require a uniform effect over strata to obtain an unconfounded summary estimate, could and should also be included.

Where does effect-measure modification fit into an analysis? Pooling assumes that the effect parameter is uniform over the strata. If it is not, then the primary interest in the analysis may shift to the description of how the effect changes by level of the stratification variable. In that case the results may be reported separately by each stratum. Doing so precludes a problem from confounding (except within strata) because the crude estimate, which is where the confounding occurs, is not presented. Thus, a stratified analysis typically aims either to evaluate and control confounding or to describe effect-measure modification, but not both. Describing effect-measure modification by giving stratum-specific estimates of an exposure effect dispatches the question of confounding, which does not arise when looking at stratum-specific results.

For analytic issues such as these, data analysis exercises provide excellent reinforcement. These exercises can be carried out with paper, pencil, and calculator. One of the subtle teaching points relates to the ambiguity of effect-measure modification. Typically, one cannot say that age either is or is not an effect modifier (this is the reason to avoid the shorter term 'effect modification'), because the answer will depend on which measure of effect is under discussion. The ambiguity of effect-measure modification comes from the choice of effect measures with which we can describe an effect. It is critical to explain how there can be effect-measure modification with respect to, say, rate difference but not with respect to rate ratio, or vice versa. Despite the ambiguity, effect-measure modification, unlike confounding, is not something that an investigator can influence. That is why we aim to *eliminate* confounding (if we have not already prevented it in the study design) but to *describe* effect-measure modification. With these distinctions carefully drawn, students will be on their way to a clear understanding of how to cope with a stratified analysis. As usual, it is the conceptual approach and not the application of statistical formulas that needs to be emphasized in the teaching.

A straightforward extension of stratified analysis is the analysis of case-control studies in which matching has been employed in subject selection. The major teaching point is simply that an analysis of matched data could be viewed as nothing more than a stratified analysis in which one stratifies by the matching

factors. This lesson tends to get lost now that conditional logistic regression is used routinely to analyse case-control studies with individual matching. Using stratified analysis is an effective way to teach students how matching biases results in case-control studies, by selecting controls according to one or more factors that are related to exposure. Surprisingly, this selection bias can be removed in the analysis. Although it involves no new principle, matching in case-control studies is so counter-intuitive that its teaching is, nevertheless, challenging. A good teaching example is the classic paper by Johnson and Johnson (1972), which presents matched data from which students can calculate effect estimates, as well as data describing the source population from which the matched subjects were drawn. It illustrates nicely the effect of matching.

Analysis of interaction

The evaluation of interaction is complicated by the fact that it is often considered a statistical issue and treated in statistics classes. The treatment is invariably an analysis of departure from some basic model structure, usually by examining product terms in a general linear model. This interaction can be described as statistical interaction, which has the same ambiguity as the related concept of effect-measure modification. As a biologist (an identification worthy of emphasis in an epidemiology course), an epidemiologist ought to be more focused on biologic interaction. This type of interaction corresponds on a basic level to the joint participation of component causes in a causal mechanism (see 'Causation and causal inference').

The main teaching point is that biologic interaction, being a description of nature, cannot have the kind of ambiguity that depends on a choice of scale or choice of effect measure. Thus, statistical interaction is scale dependent but biologic interaction is not. One can use the causal pies to derive expressions to evaluate interaction. These expressions show the fundamental role of additivity of effects as a baseline from which interaction should be measured.

Several important teaching points emerge:

1. The usual evaluation of statistical interaction, which examines departures from additivity within a logarithmic model, such as a logistic model, corresponds to evaluating departures from a multiplicative relation. Thus, the usual approach taught in statistics classes does not apply to the evaluation of biologic interaction.
2. Instead, inferences should be tied to departures from additivity of effects, additivity being the relation (with some qualifications) that one would expect for biologically independent causes. With proper handling, logistic models and other models that involve logarithmic transformations can yield straightforward evaluations of interaction based on departures from additivity of effects.

3. Statisticians have sometimes argued that both 'additivity' and 'multiplicativity' are reasonable models, each with its areas of application: for example, in multistage models, factors acting at different stages would be expected to have a multiplicative relation (Siemiatycki and Thomas 1981). The flaw in the argument is that, in the evaluation of interaction, additivity is not used as a model to describe what occurs but instead is used as a definition of independence from which interactive effects are measured. In the multistage model, although a multiplicative relation is expected for factors acting at distinct stages, it is the departure from the additivity of these effects that we use to measure the interaction.
4. In stratified analysis of ratio measures of effect that rely on pooling, one assumes that the ratio is constant over strata. This assumption is equivalent to assuming a multiplicative relation between the exposure and the stratification variable. Thus, the assumption underlying a pooled analysis is one of interaction between the exposure and the stratification variable.

The key concept to grasp in the evaluation of interaction is the importance of using a common referent category for all effect estimates, defined on the basis of joint combinations of the two exposure categories. This concept can be taught effectively using examples of stratified data. The trend towards conducting multivariate analyses in lieu of stratification has often obscured the problems of an interaction evaluation.

In theory, as more is learnt about causes of disease, there will be a greater emphasis on learning about interactions. In practice, this theory seems to have been borne out, as more interaction evaluations see the light of day. No doubt we can look forward to learning about many gene–environment interactions in years to come. Nevertheless, a precise evaluation of interaction requires so much more data than an evaluation of a single effect that most interaction evaluations are doomed to remain uninformative. As a result, good teaching examples are hard to find.

Multivariable analysis

Because students will face entire courses on multivariable analysis, it is useful to offer some counterweight. Multivariable analysis is helpful to achieve various analytic ends but its problems are often underestimated. Furthermore, there are multiple purposes for multivariable modelling, and the methods for model construction and inference differ depending on the purpose. These issues are glossed over surprisingly often in courses that deal with multivariable modelling.

To balance these influences, one should emphasize the importance of conducting stratified analysis as a primary analysis. Even in the face of many confounders, the inferences drawn from a well-conducted stratified analysis will

seldom be modified by a multivariable analysis. The approach to multivariable analysis is conceptually similar to that of stratified analysis, from the evaluation of confounding to the evaluation of interaction. Teachers should convey caution with respect to multivariable models, which have more pitfalls than stratified analysis but have become fashionable, presumably because the technology allows such analyses to be conducted more readily.

Many teachers employ computer labs and give students datasets to conduct logistic regression analyses, or Poisson regression or other multivariable modelling exercises. We suggest that the epidemiology teacher leave these computer labs to the statistics courses and place the emphasis instead on pencil-and-paper analyses of simple stratified data. The power of thought is much more potent an analytic tool than the power of computation.

Analysis of multilevel or continuous exposures

The evaluation of trend in epidemiologic data analysis is an analytic area that has been extremely weak. For decades, the only trend evaluation in a typical paper has been a declaration about the presence or absence of a 'statistically significant' trend. This reversion to qualitative thinking occurs even in papers that were much more careful to quantify effects in the analyses based on dichotomized exposures.

The main teaching point is to maintain the quantitative outlook in the evaluation of trend. One approach is simply to estimate trend rather than to evaluate it by significance tests. Reporting slope coefficients can accomplish this goal. The drawback of this approach is that the shape of the trend curve is determined by the parametric form of the statistical model used. Therefore, it is worth considering less restrictive alternatives. These alternatives are typically graphical presentations, such as spline regression or other smoothing methods that depict trend curves that follow the pattern of the actual data points, rather than a fully parametric model (Greenland 1995). These messages are easily conveyed by appropriate illustrations and some teaching examples.

Conclusion

The above discussion is merely an outline of the main teaching points and methods for a first course in epidemiologic methods. The choice and construction of the exercises and teaching examples, as well as choices about specific didactic methods, such as seminars, laboratories or workshop sessions, class discussions, and so forth, will bear heavily on the success or failure of the course. We hope that these suggestions serve to point the teacher who faces this daunting task in a promising direction.

References

Crombie, I. K. and Tomenson, J. (1981) Detection bias in endometrial cancer. *Lancet*, **2**: 308–9.

Feyerabend, P. (1975) *Against Method: Outline of an Anarchistic Theory of Knowledge*. Atlantic Highlands, NJ: Humanities Press.

Flanders, W. D. (2006) On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *European Journal of Epidemiology*, **21**: 847–53.

Greenland S. (1995) Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*, **6**: 356–65.

Hernán, M. A., Clayton, D., and Keiding, N. (2011) The Simpson's paradox unraveled. *International Journal of Epidemiology*, **40**: 780–5.

Horwitz, R. I., Feinstein, A. R., Horwitz, S. M., and Robboy, S. J. (1981) Necropsy diagnosis of endometrial cancer and detection-bias in case-control studies. *Lancet*, **2**: 66–8.

Hulka, B. S., Grimson, R. C., and Greenberg, B. G. (1981) Endometrial cancer and detection bias. *Lancet*, **2**: 817.

Johnson, S. K. and Johnson, R. E. (1972) Tonsillectomy history in Hodgkin's disease. *New England Journal of Medicine*, **287**: 1122–5.

Lang, J., Rothman, K. J., and Cann, C. I. (1998) That confounded P-value. *Epidemiology*, **9**: 7–8.

Merletti, F. and Cole, P. (1981) Detection bias and endometrial cancer. *Lancet*, **2**: 579–80.

Morgenstern, H., Kleinbaum, D. G., and Kupper, L. L. (1980) Measures of disease incidence used in epidemiologic research. *International Journal of Epidemiology*, **9**: 97–104.

Poole, C. (1987) Beyond the confidence interval. *American Journal of Public Health*, **77**: 195–9.

Reintjes, R., de Boer, A., van Pelt, W., Mintjes-de Groot, J. (2000) Simpson's paradox: an example from hospital epidemiology. *Epidemiology*, **11**: 81–3.

Rothman, K. J. (1976) Causes. *American Journal of Epidemiology*, **104**: 587–92.

Rothman, K. J., Gallacher, J., and Hatch, E. E. (2013) Why representativeness should be avoided. *International Journal of Epidemiology*, **42**: 1012–14.

Rothman, K. J., Greenland S., and Lash, T. L. (2008) *Modern Epidemiology* (3rd edn). New York: Lippincott Williams & Wilkins.

Siemiatycki, J. and Thomas, D. C. (1981) Biological models and statistical interactions: an example from multistage carcinogenesis. *International Journal of Epidemiology*, **10**: 383–7.

Thompson, R. S., Rivara, F. P., and Thompson, D. C. (1989) A case-control study of the effectiveness of bicycle safety helmets. *New England Journal of Medicine*, **320**: 1361–7.

VanderWeele, T. J. and Robins, J. M. (2007) The identification of synergism in the sufficient-component-cause framework. *Epidemiology*, **18**: 329–39.

Part 2

Exposure-oriented epidemiology

Chapter 6

Questionnaires in epidemiology

Jakob Bue Bjørner

Introduction to questionnaires in epidemiology

Use of questionnaires is an essential epidemiological tool. Epidemiological findings are often based on responses to questionnaires, which are used extensively for collecting information on exposures, outcomes, modifiers, mediators, and confounders. In such studies high quality questionnaire data are a prerequisite for drawing valid conclusions. Questionnaires should be designed and used so that they are acceptable to all participants, the response rates are high, and the responses have maximal validity (absence of systematic error/information bias) and reliability (minimal random error).

The design and use of questionnaires requires language skills (e.g. in writing questions that are easy to understand and unambiguously phrased) and some theory (e.g. a basic understanding of the cognitive processes of a person who responds to questionnaire items). Teaching must therefore combine practical training and theoretical topics. I recommend providing an initial broad overview of the topics pertaining to questionnaires in order to provide background and perspective, and thereafter giving a step-by-step introduction to each of the practical tasks of questionnaire development, validation, and use.

Question writing is particularly suited for practical training during a course. Students should first be shown examples of good and poor questionnaires (see 'Sources for standard questionnaires'). They should then write a short questionnaire (on simple issues like smoking, visits to the GP, etc.) and they should comment on each other's drafts.

Students should exercise their question-writing skills in order to experience how difficult it is to ask good questions and to learn to recognize good questions. However, it is very important that they do not reinvent the wheel. Standard questions and standard questionnaires should be used whenever possible but only after pilot testing in the study population in question. Validity and reliability are not fixed properties of a questionnaire but depend on the match between the questionnaire, the study purpose, and the respondent. Students should recognize that questionnaires used for self-administration need to be simpler than questionnaires used in an interview. As a principle, questionnaires for self-administration

need to be short and without complicated branching. They must be easy to fill out and appear well designed, with a professional finish.

Teaching objectives

1. Students should know the basics of questionnaire design (specifying a variable list, writing questions, designing response choices, and planning the coding of the answers), and they should be able to apply these principles in writing simple questionnaires on their own. They should be aware of the need for testing and revising the questionnaire several times.
2. Students should know the main data collecting methods that use questionnaires (e.g. postal surveys, phone interviews, face-to-face interviews, computer- or internet-based surveys, surveys by tablets, or cell phones), they should be able to evaluate the strengths and weaknesses of each data collection method for a particular study, and they should be able to design the questionnaire taking the chosen data collection method into account.
3. The students should know important sources for standard questionnaires and be able to perform an effective search for standard questionnaires for a given research topic.
4. Students should know basic approaches to pilot testing a questionnaire and be able to perform such a pilot test.
5. Students should know how questionnaire data are to be processed (e.g. coding, keypunching, data cleaning) and be able to design the questionnaire and the data collection to enable data processing.
6. The students should be aware that cross-cultural issues may play a role in questionnaire administration and data interpretation.
7. Finally, students need an understanding of basic psychometric concepts (e.g. reliability, validity, discrimination, sensitivity, and specificity) and how such measurement properties may be affected by questionnaire design and by study design (e.g. differential recall bias in case-control studies).

Students should also realize that designing questionnaires covering new grounds is a job that requires experience and skills that goes beyond what they learn in a single course. Besides using validated questionnaires as much as possible they should also seek professional advice.

Teaching contents

Overview

Initially, examples of questionnaires could be given to illustrate different formats of questionnaires, show examples of good questionnaires, define the concept of

an item (a question and its response options), and introduce the different data collection methods that make use of questionnaires (e.g. paper-and-pencil survey, personal interviews, phone interviews, computerized surveys). Examples of questionnaires could come from well-known epidemiological studies or from other work with which the teacher is familiar. Some sources of general surveys are the Division of Cancer Epidemiology and Genetics of the US National Cancer Institute (<http://dceg.cancer.gov/tools/design/questionnaires>) and the Epidemiology Branch of the National Institute of Environmental Health Services (<http://www.niehs.nih.gov/research/atniehs/labs/epi/questionnaires/>). Further, a variety of sources exists for questionnaires about health status (Bowling 2001, 2004; McDowell 2006; Ware et al. 2007). The latter source (the manual of the SF-36 questionnaire (Ware et al. 2007)) also illustrates how the same items can be changed in layout to suit different data collection methods. A famous example of an unfortunate design of a question is provided in Figs. 6.1 and 6.2.

A general introduction to the methods of questionnaire research could cover the following topics: questionnaire design and the properties of high quality questionnaires, methods of data collection, important sources of random and

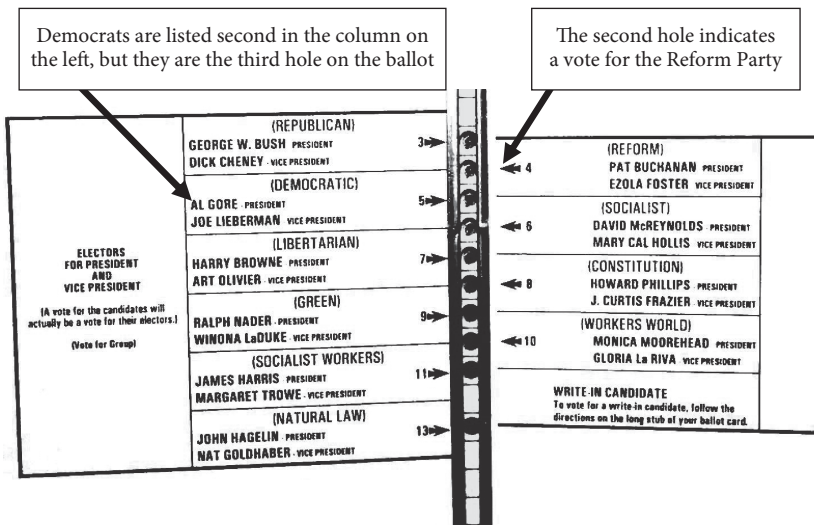


Fig. 6.1 The butterfly ballot used in Palm Beach County, Florida, for the 2000 US presidential election. The two-column layout represents a potentially confusing design (Sinclair et al. 2000). Voters who read both columns of the ballot will generally be able to identify the correct hole to punch. However, voters who only read until they find their preferred candidate may select the wrong hole.

Source: data from Robert C. Sinclair et al., Psychology: An electoral butterfly effect, *Nature*, Volume 408, Number 6813, pp.665–666, Copyright © 2000 Nature Publishing Group, Macmillan Publishers Limited, DOI:10.1038/35047160.

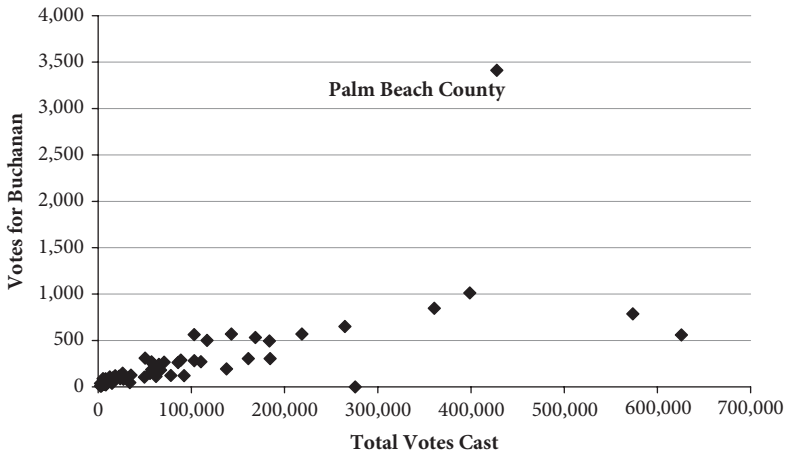


Fig. 6.2 Presidential election results for 2000 in Florida. Votes for Pat Buchanan (the Reform Party) by county: Pat Buchanan received 3,412 votes out of a total of 427,622 votes cast.

Data from *The Geocommunity website*, Copyright© 1995–2014 MindSites Group, available from www.GeoComm.com

systematic error (and the concepts of reliability and validity), data processing, and issues in data analysis.

Properties of good questionnaires

The fields of questionnaire research and psychometrics use a bewildering number of terms to describe the properties of good questionnaires. Although students need to know this terminology, it is best to focus on the relations between a moderate number of well-defined concepts that are introduced gradually (see Ware et al. 2007) for short definitions of psychometric concepts). Basically, the concepts deal with general measurement properties: that you measure what you intend to measure and that your measuring instrument is as independent as possible of the situation in which it is used. The initial discussion should focus on criteria that can be evaluated by looking at the questionnaire. The concept of *content validity* could be introduced with a discussion of whether the content of the items in a questionnaire match the variables we want to measure. Further, standard criteria for good item writing should be discussed. The items should be (Stone 1993; Streiner and Norman 2003)

1. appropriate for the group to be studied;
2. intelligible (this implies using a reading level that all respondents can understand, avoiding double negatives, using common language, and avoiding jargon);

3. concerned with only one topic for each item;
4. unambiguous (i.e. the researcher and all the respondents should understand the questions the same way regardless of their age, sex ethnic, or social background);
5. without value-laden words;
6. omnicompetent (i.e. capable of coping with all possible experiences of the respondents regarding the topic) and, when closed-form response choices are used, the response choices should be mutually exclusive and exhaustive. The respondent should always be able to find an appropriate answer (options like 'don't know' and 'not relevant' have their (limited) use); and
7. questionnaires aimed at administration by paper-and-pencil, computer, internet, PDA, tablet, or smartphone should have appropriate visual design (large enough fonts, simple and not too compact visual form, a professional and serious appearance, and response options placed either all horizontally or all vertically).

Questionnaire design

The process of designing a questionnaire could be taught in a step-by-step manner. Item development should be guided by a *variable list*, which should be kept as short as possible by excluding the items for which no clear use in the data analysis is foreseen. It should be emphasized that the reuse of relevant items from previous studies is preferable to developing items anew but in some cases writing new items is necessary.

General survey research distinguishes between measurement of *attitudes* and measurement of *behaviour* (see e.g. Sudman and Bradburn 1982). In epidemiology, *attitudes* are all the things that can only be assessed through information from the respondent (symptoms, self-rated health, personality traits, etc.), while *behaviour* could also be assessed by somebody else (number of cigarettes smoked, occupational exposures, etc.). The distinction has implications for the ways the question can be phrased, the possible response options, and the methods that can be used to evaluate the validity of the answers.

Although the response options are an integral part of the item, the topic warrants special discussion. For questions about quantities, the researcher can choose to have the respondent write the specific number or amount (e.g. *When were you born? (Date/Month/Year)* __/__/__) or offer a limited number of response choices. Although there is an advantage in getting as high precision as possible (information can always be collapsed in the analysis but not expanded), it is not advisable to require the respondent to provide greater precision than he is capable of. If this is done, the respondent will make his own guess, since there is a tendency to respond even if the answer is unknown. Such guessing produces

an unknown source of measurement error. For some pieces of information, when exact quantification is difficult, it is far easier for the respondent to choose among a limited number of response options. For the 'behaviour' type of items, typical response options are confirmation (*yes/no*), nominal categories (e.g. different types of jobs), amount, time, or frequency. For the 'attitude' kind of items, additional response types are true/false (e.g. with four categories: *definitely true*, *mostly true*, *mostly false*, and *definitely false*), agree/disagree, evaluation (e.g. *in excellent health/in poor health*, and *to a large extent/to a little extent*). For yes/no and true/false type questions, double negatives between the question and the response options should be avoided, since this often leads to confusion. This implies phrasing the question without negatives.

Another important issue is the order of items. Start with neutral questions and place sensitive questions at the end of the questionnaire. If the questionnaire has global questions (e.g. *How would you rate your health in general?*) and very specific questions (e.g. *How much pain in your right shoulder have you had during the past four weeks?*) about the same topic, it is advisable to put the global questions first, in order to avoid framing effects (undue influence of the specific items on the global items).

Finally, teaching should address how to best enable the coding of responses and the data processing.

A number of papers and text books have dealt with questionnaire design; from papers in medical journals (Stone 1993) to short textbooks (Converse and Presser 1986; Kelsey et al. 1996) and elaborate textbooks (Tourangeau et al. 2000; Streiner and Norman 2003; Dillman 2007). I recommend the short paper by Stone (1993), and this paper could be supplemented by one or more of the textbooks, depending of the length of the course. A useful comparison of advantages and disadvantages in using paper-and-pencil questionnaires, internet data collection, and interviews is given in the book by White et al. (2008).

At this point, it is helpful to introduce an individual exercise in item writing to be discussed thereafter in small groups. Here, as in all situations where students criticize each other's work, it is helpful to lay out some ground rules, emphasizing that criticism should always be constructive and include suggestions for improvement.

Sources for standard questionnaires

Since the use of standardized items should be strongly encouraged (Olsen 1998; Rosen and Olsen 2006), it is important to locate sources for items. For the measurement of health status (e.g. symptoms of disease), handbooks incorporating a large number of questionnaires are available (Bowling 2001, 2004; McDowell 2006), but often the search is less straightforward. A typical strategy is to locate

the most important studies for the given topic and then contact the authors. Some online sources are available (see 'Overview'). Most standard questionnaires are free to use but some are restricted.

Cognitive and linguistic research on questionnaires

During the last thirty years, psychologists have studied the cognitive processes in answering questionnaires. Items in a questionnaire are not simply answered by retrieving the relevant information from memory. More complex processes are involved, such as question interpretation, retrieval of partial information from memory, and construction of answers based on inference from this partial information (Bradburn et al. 1987; Tourangeau et al. 2000). During this process the present situation affects what is being 'remembered' about the past, which is one reason for recall bias. In a study of the risk of spontaneous abortion among hospital personnel, researchers compared questionnaire data and hospital records and showed that among women exposed to anaesthetic gases, 100 per cent of all miscarriages were reported, while among unexposed women only 70 per cent of all miscarriages were reported (Axelsson and Rylander 1982; see also Hennekens and Buring 1987).

Some results from cognitive research have direct rule-of-thumb implications for item writing (see e.g. Bradburn et al. 1987 and Converse et al. 1986) and suggest new methods for pilot testing (Jobe and Mingay 1990; Collins 2003; Jobe 2003; see below).

Pretesting and pilot testing of questionnaires

It is important to stress that questionnaire development involves extensive testing and revisions. The first rounds of testing should be performed by the researcher himself, by colleagues, and by supervisors. When the researcher considers the questionnaire to be without obvious flaws, the time has come for real pilot testing (Converse et al. 1986; DeMaio and Rothgeb 1996; Collins 2003). A simple scheme is to have respondents answer the questionnaire and then interview them about the answering process: were the questions appropriate, were they understandable, and was the questionnaire too long? The understanding of a sample of items should be checked and for a few items one should ask directly about the reasons for giving the specific answer (e.g. *What went through your head when you saw that question?* or *When you gave this answer, what did you think of?*). Although such interviews provide important knowledge, results and statements should be interpreted cautiously. Respondents can be fairly uncritical (Converse et al. 1986) and the items preferred by the respondents are not always the items that gives the most correct and useful information (Bradburn et al. 1987).

More elaborate techniques involve, for example, think-aloud sessions where the respondent verbalizes his/her thought process (Jobe and Mingay 1990; DeMaio and Rothgeb 1996; Collins 2003; Jobe 2003).

At this point, the students could try to conduct interviews or think-aloud sessions in small groups.

If a sufficient number of people have answered the questionnaire during pilot testing, preliminary quantitative studies can be performed. Such analyses should focus on the magnitude of non-response, on items where all (or nearly all) responses are in one category (indicating that the response options are not targeted at the population of interest), and on obvious response errors.

Data collection

Students should learn about the main methods of data collection (postal surveys, phone interviews, face-to-face interviews, computer and internet surveys), the advantages and disadvantages of each method, and the implications of the data collection method for questionnaire design (for good reviews see Dillman 2007 and Kelsey et al. 1996).

General rules for conducting structured interviews should be outlined. The ways to conduct follow-up to enhance response rates should be discussed as well as the problems related to non-responders and non-responses for part of the questionnaire. Discuss also the possibilities of characterizing non-responders on a few important variables in order to evaluate the likely effects of non-response. When the data collection involves collaboration with others (e.g. in distributing questionnaires), it is important to stress mutual interest and respect are essential for the success of the study.

Data processing

Data processing involves the coding of responses, keypunching or scanning, quality control, and data cleaning. It is important to discuss principles for coding (see e.g. Ware et al. 2007). It should be emphasized that the coding strategy must be established before data are collected.

When conducting phone interviews or computerized assessment, data processing may be performed simultaneously using computer programs with pre-programmed data entry screens and built-in quality controls (e.g. out of range responses not allowed). If response codes are keypunched from coded questionnaires directly into ASCII files, it is advisable to conduct the keypunching twice and recheck all cases of discrepancy.

At this point, the students should carry out short exercises in data processing.

Cross-cultural issues

Cross-cultural issues are of increasing importance in questionnaire research. Most standard questionnaires have been written in English, and the questionnaire has to be translated and adapted when used for respondents with other native languages. A number of steps have to be taken to ensure comparability between the different versions if results are to be generalized across countries (see Guillemin et al. 1993; Bullinger et al. 1998). In general, questions should be phrased in the respondents' native language, and interviewers should never be asked to translate from a questionnaire written for a different language.

Students at the postgraduate level should have some knowledge of basic methods for testing cross-language equivalence: independent ratings, forwards-backwards translation, and psychometric methods (see 'Basic psychometric concepts'). It should be stressed that sometimes institutional or cultural differences can make it impossible to construct questionnaires that are exactly equivalent (e.g. different ways of organizing the educational system makes it difficult to compare education across some countries, pregnancy planning may mean something different in different cultures, and words like 'wheezing', for example, do not exist in all languages).

Basic psychometric concepts

The field of psychometrics is complex and I recommend starting with simple concepts and using examples. Initially, issues like data completeness (low frequency of non-response to each item), item discrimination, and floor and ceiling effects should be discussed (see e.g. Streiner and Norman 2003; Ware et al. 2007).

Psychometric analysis distinguishes between random measurement error and systematic (non-random) measurement error. The term 'latent variable' is used for the measurement we would have achieved if we had been able to measure without error at all ('true measure'). For measures that are on a continuous scale (or can at least take a large number of ordered values) and where the latent variable is also conceptualized as continuous, random measurement error is described through the concept of reliability, the proportion of the observed score variance that is explained by the latent variable. Thus, high reliability means low random measurement error. The concept of validity is used to characterize lack of systematic measurement error.

A common way to deal with random measurement error and floor and ceiling problems of individual items is to ask several questions about the same topic

and combine the answers to these questions (e.g. by taking the average of the item responses or using the simple sum of item responses). Such combination of items into a scale assumes unidimensionality, that is, that all items measure the same thing. The appropriateness of such scale construction is tested through latent structure analysis by using methods like factor analysis, structural equation modelling (Muthén and Muthén 2004; Skrondal and Rabe-Hesketh 2004), item-response analysis (van der Linden and Hambleton 1997; Edelen and Reeve 2007), and latent class analysis. While these techniques go beyond standard introduction to questionnaires, a few basic concepts and problems may be mentioned. One such issue is the importance of distinguishing between effect indicator models and cause indicator models (Bollen and Lennox 1991). Latent structure analysis has been developed in areas where the items are naturally conceptualized as effects of a latent cause (like intelligence, anxiety, and depression). However, in many areas of epidemiology, the items must be seen as causes of concept assessed by a composite scale (e.g. major life events are combined to a scale not because of their common cause but because they are expected to lead to the same effects). While the effect indicator model can be analysed by straightforward application of latent structure analysis (and the reliability can be assessed with techniques like Cronbach's alpha), standard application of these techniques is wrong when the items are causal indicators (Bollen and Lennox 1991).

Another issue in latent structure analysis is to use techniques that respect the measurement level of the observed variables. While questionnaires often use ordinal response scales with two to seven levels, the so-called classical psychometrical methods (Carmines and Zeller 1979) assumes that the observed variables are continuous. However, if the items are skewed or have few response categories, treating them as continuous may cause errors. In such situations item-response theory methodology (Edelen and Reeve 2007; van der Linden and Hambleton 1997) or structural equation modelling of categorical data (Muthén and Muthén 2004; Skrondal and Rabe-Hesketh 2004) should be used.

Over the past few years, epidemiology has witnessed an increased interest in analysis of measurement error through graphical models (Hernán and Cole 2009; le Cessie et al. 2012; VanderWeele and Hernán 2012) and in statistical methods for correction of measurement error (Rosner et al. 1990; Cole et al. 2006; Messer and Natarajan 2008). While these techniques can be described at postgraduate level, it is important to emphasize that they rely on strong assumption about the measurement error model. Reduction of measurement error—through the best possible questionnaire as well as by other aspects of study design—is more important than statistical corrections.

Teaching method and format

I suggest that undergraduate students are taught the basic principles in at least two lectures. At the postgraduate level, for example, in PhD courses, it is our experience that much more time should be spent on the topic. A one- to two-day course is the minimum, and students who work with measuring latent concepts need at least a one-week course.

For a relatively short course (two to four lectures), classroom teaching with some individual and group exercises is preferable. For a longer course, students could be assigned to do a small questionnaire study in groups of three to five persons. Such a study should take them through all the steps of variable list construction, reviews of existing questionnaires, questionnaire design, and data collection, processing, analysis, and presentation.

Assessing students' achievement

Since the questionnaire topic requires both practical skills and theoretical knowledge, students are best assessed through a combination of item writing and examination.

References

Axelsson, G. and Rylander, R. (1982) Exposure to anaesthetic gases and spontaneous abortion: response bias in a postal questionnaire study. *International Journal of Epidemiology*, **11**: 250–6.

Bollen, K. A. and Lennox, R. (1991) Conventional wisdom on measurement: a structural equation perspective. *Psychological Bulletin*, **110**: 305–14.

Bowling, A. (2001) *Measuring Disease—A Review of Disease-Specific Quality of Life Measurement Scales*. Buckingham: Open University Press.

Bowling, A. (2004) *Measuring Health—A Review of Quality of Life Measurement Scales*. Buckingham: Open University Press.

Bradburn, N. M., Rips, L. J., and Shevell, S. K. (1987) Answering autobiographical questions: the impact of memory and inference on surveys. *Science*, **236**: 157–61.

Bullinger, M. et al. (1998) Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. *Journal of Clinical Epidemiology*, **51**: 913–23.

Carmines, E. G. and Zeller, R. A. (1979) *Reliability and Validity Assessment*. Beverly Hills: Sage Publications.

Cole, S. R., Chu, H., and Greenland, S. (2006) Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*, **35**: 1074–81.

Collins, D. (2003) Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research*, **12**: 229–38.

Converse, J. M. and Presser, S. (1986) *Survey Questions—Handcrafting the Standardized Questionnaire*. London: Sage Publications.

DeMaio, T. J. and Rothgeb, J. M. (1996) 'Cognitive interviewing techniques: in the lab and in the field', in N. Schwarz and S. Sudman, eds, *Answering Questions. Methodology for*

Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass, pp. 177–98.

Dillman, D. (2007) *Mail and Internet Surveys: The Tailored Design Method—2007 Update with New Internet, Visual, and Mixed-Mode Guide*. New York: John Wiley & Sons.

Edelen, M. O. and Reeve, B. B. (2007) Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, **16** Suppl. 1: 5–18.

Guillemin, F., Bombardier, C., and Beaton, D. (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*, **46**: 1417–32.

Hennekens, C. H. and Buring, J. E. (1987) *Epidemiology in Medicine*. Boston, MA: Little, Brown and Company.

Hernán, M. A. and Cole, S. R. (2009) Invited commentary: causal diagrams and measurement bias. *American Journal of Epidemiology*, **170**: 959–62.

Jobe, J. B. (2003) Cognitive psychology and self-reports: models and methods. *Quality of Life Research*, **12**: 219–27.

Jobe, J. B. and Mingay, D. J. (1990) Cognitive laboratory approach to designing questionnaires for surveys of the elderly. *Public Health Report*, **105**: 518–24.

Kelsey, J. L., Whittemore, A. S., Evans, A. S., and Thompson, W. D. (1996) 'Measurement I: Questionnaires', in J. L. Kelsey, ed., *Methods in Observational Epidemiology*. Oxford: Oxford University Press, pp. 364–90.

le Cessie, S., Debeij, J., Rosendaal, F. R., Cannegieter, S. C., and Vandenbroucke, J. P. (2012) Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology*, **23**: 551–60.

McDowell, I. (2006) *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford: Oxford University Press.

Messer, K. and Natarajan, L. (2008) Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. *Statistics in Medicine*, **27**: 6332–50.

Muthén, B. O. and Muthén, L. (2004) *Mplus User's Guide, Version 3*. Los Angeles, Muthén & Muthén.

Olsen, J. (1998) Epidemiology deserves better questionnaires. IEA European Questionnaire Group. International Epidemiological Association. *International Journal of Epidemiology*, **27**: 935.

Rosen, T. and Olsen, J. (2006) Invited commentary: the art of making questionnaires better. *American Journal of Epidemiology*, **164**: 1145–9.

Rosner, B., Spiegelman, D., and Willett, W. C. (1990) Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*, **132**: 734–45.

Sinclair, R. C., Mark, M. M., Moore, S. E., Lavis, C. A., and Soldat, A. S. (2000) An electoral butterfly effect. *Nature*, **408**: 665–6.

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Stone, D. H. (1993) Design a questionnaire. *British Medical Journal*, **307**: 1264–6.

Streiner, D. L. and Norman, G. R. (2003) *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford: Oxford University Press.

Sudman, S. and Bradburn, N. M. (1982) *Asking Questions. A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

van der Linden, W. J. and Hambleton, R. K. (1997) *Handbook of Modern Item Response Theory*. Berlin: Springer.

VanderWeele, T. J. and Hernán, M. A. (2012) Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. *American Journal of Epidemiology*, **175**: 1303–10.

Ware, J. E. Jr., Kosinski, M., Bjorner, J. B., Turner-Bowker, D. M., and Maruish, M. (2007) *User's Manual for the SF-36v2(tm) Health Survey*. Lincoln, RI: QualityMetric Inc.

White, E., Armstrong, B. K., and Saracci, R. (2008) *Principles of Exposure Measurement in Epidemiology: Collecting, Evaluating, and Improving Measures of Disease Risk Factors*. Oxford: Oxford University Press.

Chapter 7

Environment

Anders Ahlbom

Introduction to environmental epidemiology

Environment epidemiology in this chapter is taken to mean the study of associations between exposure to chemical and physical factors in the general environment and risks of disease or ill health. There is strong and increasing interest in this, as recently evidenced by the quality and size of the 2013 Society for Environmental Epidemiology conference. We currently face a number of environmental health issues which will depend on epidemiology for their resolution (e.g. air pollution, noise, hormone disruptors, allergy). Global warming and climate changes create new challenges for environmental epidemiology. Interest is also fuelled by the fact that considerable advancements currently are being made in many of the areas in environmental epidemiology. One reason for this is that many well-planned, long-term prospective studies have started to yield results. Several of those have applied modern molecular techniques to facilitate the study of interactions between genes and environmental factors and thereby making it possible to identify sensitive subpopulations on which environmental factors may have a particularly strong effect. In many instances these findings have also contributed to the elucidation of mechanisms behind environmental health effects.

The general principles of environmental epidemiology are of course no different from the principles of other branches of epidemiology. Yet, teaching environmental epidemiology may require certain considerations, besides the obvious, that examples should be taken from the environmental area. All applied areas of epidemiology require joint understanding of both epidemiology and the key features of the subject matter area, and epidemiology is only one of the scientific disciplines that provide data for environmental health research. Thus, many of those who study environmental epidemiology do not have a primary background in epidemiology but rather in other areas of relevance to environmental health, such as toxicology, biochemistry, or engineering and of course medicine.

From the teaching perspective it may be a good idea to consider that someone with a basic training in an experimental science may feel uncomfortable about the

observational nature of most epidemiology. This may affect the very basic issue of assessing causality. It is not uncommon that students think of epidemiology as simply providing statistical associations and that experimental research is required to unveil mechanisms and prove causality. It is therefore essential to have a comprehensive discussion of the meaning of causality and of the principles of evaluating causality. That discussion will include the need to combine epidemiologic data with biological plausibility and experimental data. It is still common that people without basic epidemiology training refer to the Hill criteria (Hill 1965) as the key reference for evaluating causality. While the Hill criteria were revolutionary when they were published and still make sense in many ways, they are outdated in other respects, and most current textbooks have updated and more comprehensive and systematic discussions about principles for evaluation of causality.

The evaluation of confounding is related to this topic because confounding may appear as the main alternative to causality to someone used to randomization. Confounding may be defined as the presence of a baseline difference in risk between exposed and unexposed subjects that exists regardless of whether the exposure has an effect on risk. For confounding to occur, two associations must be in place simultaneously, namely, between the confounder and the disease and between the confounder and the exposure. It is important to point out for students that, in order to assess the possibility that an observed association is due to confounding rather than to causation, one must keep the magnitude of the associations in mind. The magnitude of a confounding effect depends on the strength of the two involved associations. It has been shown that, for a confounder to have a noteworthy effect, both the involved associations must be strong. Generally, one cannot expect confounding to create a relative risk exceeding, say, 1.5 or 2.0 and, even so, both associations would have to be quite strong. This is rather useful information for students to possess.

The epidemiologist's approach to exposure assessment may also be alien to an experimental scientist because it may appear to lack precision for someone used to administering internal doses of radiation or a chemical substance to cell cultures. While the epidemiologist certainly appreciates and demands valid exposure assessment, there are levels of detail and precision that an epidemiologic study cannot take advantage of, while other aspects such as the actual time period and duration of exposure to study subjects may be crucial. It may be useful for the teacher to use the triad of true dose/exposure, operational/practical definition of exposure, and measurement of exposure. The teacher could then go on and explain that the exposure measurements that are used in any study, epidemiologic or experimental, are proxies or estimations of the true dose/exposure.

A related topic that needs to be discussed is that levels of exposure to environmental agents often are low, for example, compared to levels that may occur

in occupational settings. The implication is not that environmental factors are less important, because the number of exposed people may still be large. However, health effects may be difficult to observe, and this requires special attention in the design of studies. One crucial aspect is accuracy in measurement of exposure, as discussed above. Another is the possibility to study particularly sensitive groups of people, as also discussed above. Study size is of course also important, because causality may be more difficult to separate from random variability in situations with weak effects.

'Cocktail exposures' or 'cocktail effects' are often mentioned in environmental health sciences. This refers to the possibility that a plenitude of exposures may occur simultaneously and that the combination of exposures could be particularly hazardous. From the teaching point of view this leads into a discussion of whether the ingredients in the cocktail have independent effects on health that are just added when exposed to the cocktail, or if the ingredients interact and amplify each other's effects. A discussion of interaction between causes of disease is required to clarify this.

A rather different, yet important, aspect to keep in mind when teaching environmental epidemiology is that decision-makers in various capacities need information about environmental health risks. Their needs are somewhat different from the scientists and it will not be sufficient to address issues of causality to satisfy the decision-makers. They need quantitative data on magnitudes of risks and on the potential public health impact. Thus, they demand information about the risk function, that is, the magnitude of the risk for different levels of exposure and for defined subgroups of the population. They will also require data on exposure distribution in the population, which perhaps is not the epidemiologists' task. However, the decision-makers need the epidemiologist to combine data into an assessment of the potential impact of an exposure to public health in order to inform about mitigation strategies.

It was mentioned previously that all applied epidemiology needs a combination of epidemiologic methods and subject matter knowledge. It is possible that students in courses in environmental epidemiology often come not from epidemiology but from some other discipline of relevance to environmental health science. Some of the comments above are directed towards teaching needs that might arise from this. There is also the reversed situation, namely, that students with training in epidemiologic methods and no particular background in, for example, toxicology, biochemistry, or engineering take a course in environmental epidemiology. The question then arises as to what extent the course should include such elements. If the course is included in a wider teaching programme, it is natural for the programme to include a course that provides training in these and other topics that are needed for students who come,

for example, from a statistical or social science background. If the course in environmental epidemiology is a stand-alone one, it is still preferable if time could be spent on a review of the nature of data that experimental sciences can provide and what they can and cannot tell us.

The most important and difficult part to teach in a course in environmental epidemiology is perhaps how human data from epidemiologic studies are pooled with mechanistic data from experimental studies. Epidemiologic results that are not backed up by biological plausibility or findings from experimental research are difficult to interpret and will normally not lead to far-reaching conclusions. The same holds for toxicological results without corresponding findings from epidemiologic research. Both those situations will result in uncertain health risk assessments and will sometimes result in conclusions such as 'exposure to X could increase the risk of disease Y'. This leaves the room wide open for interpretation and often results in controversies and situations that are complex for society to handle.

The ideal situation is when there is comprehensive data from both epidemiologic and experimental research that point in the same direction. Overall conclusions are easy in those situations. If experimental data indicates that a mechanism exists by which exposure X can lead to disease Y and at the same time epidemiologic research finds such an association, the overall assessment is straightforward. The same holds of course when both research areas fail to find links between the exposure and the disease. It is more difficult when research findings are inconsistent.

It may seem obvious, but it is still worth stressing that, in research aimed at clarifying the association between an environmental factor and a disease risk in humans, epidemiologic data address this directly and are therefore of highest relevance. A justification for this and a principle for how this is applied may be found in the preamble of any of the International Agency for Research on Cancer monographs (International Agency for Research on Cancer 2014) on the evaluation of carcinogenic risks to humans. From in vivo and in vitro studies a generalization will always have to be made to the human situation and so external validity is always an issue. On the other hand, it is usually easier to achieve a higher internal validity in the experimental setting than in an observational epidemiologic study.

Teaching objectives

After the course the students should

1. have a good understanding of basic principles and practice in epidemiologic research;

2. have a good understanding of some methodological issues of particular relevance to environmental epidemiology;
3. appreciate the importance of pooling data from various branches of science when making health risk assessments and in particular to appreciate the role of epidemiologic data; and
4. appreciate what kind of quantitative data decision-makers need.

Teaching contents

Basic principles and practice of epidemiology

It would be ideal if students in a course on environmental epidemiology could come with a strong background in basic epidemiology. However, in practice this has turned out not always to be the case. Even when students have this background, they may be used to different concepts and principles. Therefore, some time in a course on environmental epidemiology should be spent on basic epidemiologic principles. The exception is when this course is part of a wider programme that also includes introductory courses that can be requested for admission. This introductory part of the course has probably a standard layout for a general course in epidemiology but with examples taken from the field of environment and health. Particular emphasis should be given to the concept of causality and principles for evaluating causality.

Introduction to toxicology and other experimental data

This section should provide a brief introduction to the nature of information that toxicology and some other relevant disciplines provide.

Issues of particular relevance to environmental epidemiology

The topics and study designs in environmental epidemiology cover a wide range and it is not obvious which special methodological issues should be given priority in a course. Still, there are a few issues worth special attention. This should of course be adjusted if students have interest in a particular issue or if some important study just appears that has used some special features. Topics to cover include the fact that many environmental health effects often are weak and that this requires some special attention, the fact that there is a limit to how strong effects that confounding can generate, the requirements on exposure assessment in environmental epidemiology, and the possibility of taking advantage of modern molecular biology to identify sensitive groups. The section should also have a review of the methods used to supply decision-makers with

the quantitative data required. The course could also include more specialized designs such as time trends to look at acute effects in relation to variation over time in exposure. Another common method in environmental epidemiology is the use of geographical information systems, with data on exposure or markers of exposure. Occasionally, cross-sectional studies with simultaneously collected data on exposure and health are used and it would be reasonable to discuss the difficulties attached to this approach.

Overall health risk assessment

The final session is devoted to the overall health risk assessment, in which data are evaluated and combined.

Comments

The basic principles of environmental epidemiology are of course no different than those in other branches of epidemiology. This chapter underscores some of the principles that may be of particular interest when teaching epidemiology to an audience with a background other than epidemiology. In particular the discussion has focused on some of the consequences that arise from the observational rather than experimental nature of most epidemiologic research. Other important aspects are the (sometimes) weak effects of environmental risk factors that create particular difficulties for environmental epidemiologists. The prevalence of environmental exposures may sometimes be low but there are also instances with ubiquitous exposure. Both situations create challenges for the epidemiologist. The rapidly evolving field of molecular biology will increasingly provide new tools such as biomarkers and markers for individual susceptibility and it will be essential to incorporate these into study designs. The most important characteristic for environmental epidemiology is perhaps the need for good collaboration with experts in related areas, often for the purpose of exposure characterization but of course also for mechanistic understanding and the formulation of hypotheses.

Teaching method and format

The teaching format I prefer is a combination of lectures, exercises, and case studies. Each topic would first be presented in a lecture. It would then be followed up with simple exercises. These are performed individually and reviewed in class at the next lecture. Each topic is also followed up with a case study, preferably in the form of a published article that is used to illustrate a particular topic. The students read the paper for themselves, together with specific questions related to the paper. They then discuss the questions in small groups and, finally, in a plenary.

Assessing students' achievements

The method for evaluation of students depends on whether the course is taken for credit or not. It also depends on the length of the course, because one does not want to use too large a proportion of the time for the test. One should try to integrate the evaluation in the teaching and my preference is a series of quizzes. They may take only fifteen minutes to do in class. Discussion of the quizzes afterwards may be an instructive way of reviewing certain topics.

References

Hill, B. A. (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **58**: 295–300.

International Agency for Research on Cancer. (2014) *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. <http://monographs.iarc.fr/ENG/Classification/>, accessed 4 November 2014.

Chapter 8

Occupational epidemiology

Neil Pearce

Introduction to occupational epidemiology

Occupational epidemiology is the study of the distribution and causes of illness and injury that result from hazardous workplace exposures (Checkoway et al. 2004). Concerns about adverse health consequences of occupational exposures date back to Hippocrates' warnings to physicians to explore patients' environmental lifestyle and vocational backgrounds. In 1700 the Italian physician Ramazzini described numerous occupational diseases such as silicosis among stonemasons, ocular disorders among glass blowers, and neurological toxicity among tradesmen exposed to mercury. Other early descriptions of occupational diseases reported pneumoconiosis among miners of gold and silver in Germany and the former Czechoslovakia, and the identification of soot as the cause of scrotal cancer in London chimney sweeps by Percival Pott in 1775 (Checkoway et al. 2004).

More recent examples include the recognition of asbestos-associated disease including asbestosis, lung cancer, and mesothelioma, as well as papers reporting associations of industrial solvents with Alzheimer's disease (Axelson 1995; Kukull et al. 1995), various metals (Gorrell et al. 1997) and pesticides with Parkinson's disease (Seidler et al. 1996; Liou et al. 1997), man-made mineral fibres with lung cancer (Boffetta et al. 1999), and polycyclic aromatic hydrocarbons with lung and bladder cancers among workers in coke ovens, aluminium smelters, and metal foundries (Boffetta et al. 1997).

Until recently, most occupational epidemiology studies have been conducted in Western countries. However, the number of workers in industries involving a risk of cancer and other occupational diseases is increasing in developing countries, partly as a result of the transfer of these industries from industrialized countries (Pearce and Matos 1994). Thus, it is increasingly important to take a global view of occupational disease, just as for health more generally (Pearce 2004). Occupational health is becoming more, rather than less, important on a

global basis (Guidotti 2011), as recent changes in the global economy have shifted manufacturing and attendant hazards out of Europe and North America to rapidly developing economies in which millions of workers could be exposed to serious risks (Pearce 2012b), an issue that, unfortunately, is omitted from most discussions on 'global public health' (Cullinan and Pearce 2011).

Teaching objectives

Students in occupational epidemiology courses may include trained epidemiologists who are learning about occupational epidemiology for the first time, occupational hygienists, and health care professionals (e.g. occupational physicians, occupational health nurses, and students of medicine and nursing). Increasingly, students will come from a wide range of countries—high-income countries (HICs) and low-and-middle-income countries (LMICs)—and it is important that courses address the methodology and substantive findings of occupational epidemiology on a global basis, rather than being primarily based on examples and methods from HICs. It is also important to distinguish between teaching occupational epidemiology to researchers and teaching epidemiology to public health officers and occupational physicians. Researchers often have a stronger interest in theoretical issues, whereas the latter groups may have a stronger interest in specific occupational health problems (Merletti and Comba 1992). In this chapter I assume that students have previously undertaken an introductory epidemiology course, and I focus on issues which are relatively unique to, or which receive greater emphasis in, occupational epidemiology. I emphasize a problem-based approach which involves learning occupational epidemiology in the context of active experience in developing, conducting, and reporting on a (hypothetical) occupational epidemiology study. By the end of the course, participants should be able to

- describe the strengths and limitations of occupational epidemiology and particularly its potential for use in solving real workplace health problems and developing preventive interventions;
- discuss the key features of occupational epidemiology that distinguish it from other fields of epidemiology;
- identify the major global occupational health problems and their potential solutions and key research questions;
- consult with managers, unions, workers, researchers, government agencies, and other interested parties in developing a research question;
- design an occupational epidemiology study that addresses this research question;

- write a research proposal suitable for submission for funding; and
- communicate the findings of research both in academic form and in the form of a press statement for the general media.

Of course, these objectives will be met to a greater or lesser extent depending on the background, experience, and ability of the course participants but the specific course content, and the level of detail and methodological rigour, can be modified accordingly.

Teaching content

In this section I discuss the key topics that should be covered as a complement to a more general introductory epidemiology course.

Reasons for conducting occupational epidemiology studies

The primary objective of occupational epidemiology is to identify hazardous work place exposures in order to prevent occupational disease (Mendes and Costa Dias 2011). The potential for prevention is often much greater than in other fields of epidemiology, since workplace exposures can often be readily identified and removed. This can be illustrated using examples of successful prevention, including non-introduction of hazards into the workplace, removal of hazards, reduction in exposure levels, reduction in hazardous activities, and increased protection (Swerdlow 1990). It is therefore valuable to discuss current workplace health and safety issues with management, unions, and workers and to identify the potential health hazards that are currently of most concern and for which there is the greatest potential for prevention.

A second objective of occupational epidemiology is to provide information that can be used in risk assessment and in the prevention of hazards in the general population. For example, it is notoriously difficult to study the health effects of occasional low-level pesticide exposure in the general population, and it is more valuable to conduct studies of heavily exposed pesticide production workers and sprayers and to use the findings from these studies to estimate the risk from lower levels of exposure in the general population. Similarly, exposure levels are generally higher in LMICs, so collaborative studies between researchers in HICs and those in LMICs may be particularly useful (Pearce 2011a).

Once again, a problem-based approach can be valuable. For example, students can be introduced to these issues by discussing a problem that is currently of community concern (e.g. the risks of environmental pesticide exposure, exposure to electromagnetic fields, asbestos exposure in the home), and by discussions with some of the interested parties in the community and in industry,

followed by consideration of other exposed populations (e.g. occupational exposures and/or exposures in other countries).

A further reason for conducting occupational epidemiology studies is to evaluate the effects of workplace interventions such as the removal of a hazardous exposure and its replacement with a substitute believed to be less hazardous. For example, asbestos has increasingly been replaced by various man-made mineral fibres, which appear to also involve an increased risk of asbestos-related diseases such as lung cancer (Boffetta et al. 1999). It is therefore important to emphasize the importance of following up on previous epidemiological studies, ascertaining what preventive measures have been adopted in response to their findings, and how effective these measures have been. For example, in a longitudinal study of occupational respiratory disease, it is relatively straightforward to assess the effects of changes in health and safety practices and provision of protective equipment over time (Slater et al. 2000).

Political and ethical issues

A problem-based approach will inevitably lead to the discussion of controversies in occupational epidemiology and risk assessment. Discussing such controversies can be valuable in demonstrating the strengths and weaknesses of occupational epidemiology, its wider role in causal assessment and risk assessment, and the political, economic, and social influences that often affect which hypotheses are chosen for study, which methods are used, and how the findings are interpreted (Pearce 2008). These issues are relevant to all epidemiological studies but are often particularly important in occupational epidemiology, as well as in other fields of epidemiology that involve vested interests such as environmental epidemiology and pharmacoepidemiology (Barnes and Bero 1998; Michaels 2005a, b; Egilman 2006; Monforton 2006; Pearce 2007a).

Unfortunately, such discussions can often involve an undue emphasis on criticism of published studies, and this is often encouraged in courses that overemphasize the role of criticism under a simplistic interpretation of the Popperian philosophy of science (Pearce 2007b). Such a one-sided emphasis on criticism can lead to disillusionment among students as to the validity and value of occupational epidemiology. Thus, it is valuable to ask students to design a study of a specific issue and to write a research protocol for it, rather than asking them merely to critique a published paper. This approach is not only invaluable in terms of practical experience with epidemiologic study design but can also teach students the practical compromises that are often involved in conducting occupational epidemiology studies. They can therefore learn that no study can be perfect or definitive and that the aim of a single study should be to contribute to the pool of information available for scientific and

public health decision-making, rather than requiring hazard or safety to be proven in a single study.

The potential for preventive action and the considerable vested interests which may hasten or delay such action mean that ethical issues play a major role in occupational epidemiology. In particular, the need to conduct further epidemiological studies should not be a reason to delay preventive measures. Furthermore, it is important that reports of occupational epidemiology studies, and criticisms of these reports, should openly acknowledge sources of funding and any other potential conflicts of interest (Davey Smith 2001). A related issue is whether occupational epidemiology studies should be 'registered' in advance to reduce the likelihood of chance findings for exposures and hypotheses that were not of a priori interest, an issue which has generated considerable debate in terms of both ethics and science (Pearce 2011b).

Methodological issues

Healthy worker effect

A major issue in occupational studies is the healthy worker effect (Checkoway et al. 2004). The typically lower relative risk of death or chronic disease in an occupational cohort occurs because relatively healthy individuals are likely to gain employment and remain employed. Thus, the initial selection occurs at time of hire in that relatively healthy persons are more likely to seek and to be offered employment; the most direct way to partially control for this phenomenon is to stratify on initial employment status; that is, to compare the asthma morbidity of a particular workforce with that of other employed persons rather than with a general population sample (which includes invalids and the unemployed). The second key aspect of the health worker effect is the selection of unhealthy persons out of the workforce. In cohort studies, this problem can be partially addressed by considering each worker's employment status (i.e. active or non-active worker) and time since first employment at a particular time and then controlling for it as a confounder (Steenland and Stayner 1991; Pearce 1992). However, the healthy worker effect may also be particularly strong in cross-sectional studies (Eisen 1995), although this can be partially diminished by focusing on exposures that occurred prior to the onset of symptoms (Eisen et al. 1997).

The healthy worker effect means that considerable attention should be given to the selection of appropriate comparison populations. Despite the limitations of standardized mortality ratio (SMR) analyses using the general population as a comparison, this is usually the most practical option for initial analyses. However, alternative comparison populations (e.g. regional mortality rates, other

employed workers) should also be considered. The value of internal comparisons, and the importance of defining a cohort so that such comparisons are feasible and meaningful, should also be emphasized.

However, although the use of an internal reference group may control for initial employment status and therefore for the initial selection into employment, it will not necessarily eliminate other forms of bias. In particular, exposure (or the termination of exposure through leaving employment) can be a cause and/or a consequence of occupational disease and can affect the subsequent risk of mortality. Such 'intermediate variables' should not be routinely controlled using standard techniques, and special techniques are required to avoid adding bias (Pearce and Greenland 2004).

Exposure data

A feature of occupational epidemiology studies is the frequent use of job exposure matrices to estimate historical exposures. The sources of data for a job exposure matrix (JEM) include industrial hygiene sampling data, process descriptions and flow charts, plant production records, inspection and accident reports, engineering documentation, and biological monitoring data (Checkoway et al. 2004). However, the starting point for the development of job exposure matrices and the estimation of individual historical exposures is the personnel records. It is therefore important to use real examples of personnel records in teaching and to discuss the practical aspects of classifying and grouping departments and job titles.

In the last two decades, there has been increasing emphasis on the use of molecular markers of internal dose. Nevertheless, there were a number of major limitations of available biomarkers of exposure (Armstrong et al. 1992), particularly with regard to historical exposures (Pearce et al. 1995). Some biomarkers are better than others in this respect (particularly markers of exposure to biological agents) but even the best markers of chemical exposures usually reflect only the last few weeks or months of exposure. Thus the use of work history records in combination with a job exposure matrix (based on historical exposure measurements of work areas rather than individuals) is often more valid than current exposure measurements (whether based on environmental measurements or biomarkers) if the aim is to estimate historical exposure levels (Checkoway et al. 2004).

However, in recent years, a broad range of 'omics' methodologies have become available, including not only genetic markers but also epigenetic markers of environmental exposures and related measures, such as proteomics and metabolomics (Kumpula et al. 2010; Jennen et al. 2011; Kamburov et al. 2011). These technologies are undergoing rapid development; in most cases they are

not yet valid or inexpensive enough to use routinely in occupational epidemiology studies but this situation is likely to change over the next decade (Wild 2005; Rappaport and Smith 2010; Vlaanderen et al. 2010; Rappaport 2011, 2012). The emphasis should be on using 'appropriate technology' and that the most appropriate approach (questionnaires, environmental measurements, or biological measurements) will vary from study to study and from exposure to exposure within the same study (or within the same complex chemical mixture (e.g. in welding fumes)). This can be illustrated by a problem-based approach in which different exposure estimation methods are shown to be appropriate for different studies (Steenland 1993).

Confounding

Confounding is of concern in occupational epidemiology, as in other fields of epidemiology (Pearce and Greenland 2004). However, to be a significant confounder, a factor must be independently strongly predictive of disease and strongly associated with exposure (Rothman and Greenland 1998). Thus, confounding is often relatively weak in occupational studies, particularly when comparing 'exposed' and 'non-exposed' manual ('blue collar') workers, since there are usually few important differences in lifestyle between different groups of workers. For example, Siemiatycki et al. (1988) found that confounding by smoking is generally very weak for internal comparisons in which exposed workers are compared with non-exposed workers in the same factory or industry. If it is not possible to obtain confounder information for any study subjects, it may still be possible to estimate how strong the confounding is likely to be from particular risk factors. This is often done in studies of occupational causes of lung cancer, where smoking is a potential confounder but smoking information is rarely available. For example, Axelson (1978) found that, for plausible estimates of the smoking prevalence in occupational populations, confounding by smoking can rarely account for a relative risk of lung cancer greater than 1.5. Course participants can repeat this exercise for other situations (e.g. by estimating the potential for confounding in their proposed study).

However, it should also be emphasized that occupational exposures often involve complex mixtures such as welding fumes. In this situation, confounding by 'external' exposures such as smoking is likely to be weak but there may be a significant 'identification problem' with regard to the etiologically relevant constituent(s) of the complex occupational exposures.

Study design

The study design options in occupational epidemiology are the same as for other fields of epidemiology (Checkoway et al. 2004) but the methodological

issues involved may differ (see above), as may the emphasis given to specific study designs. In particular, cohort studies are relatively common in occupational epidemiology since most occupational exposures are 'rare' in the general population but suitable cohorts can be identified through routine records such as personnel records or union membership records (Checkoway et al. 2004). Thus, it is valuable to review the standard study design options but to give particular attention to the specific issues and methodological characteristics of occupational studies.

Routine mortality and morbidity data

In many countries death certificates include the deceased person's current or most recent occupation and it is therefore possible to estimate national or regional death rates for specific occupations. Occupational information may also be available for hospital admissions and for registers of diseases such as cancer. Thus, analyses of existing data sources play an important role in occupational epidemiology ('t Mannetje and Pearce 2010). Direct age-standardization and other methods for 'descriptive' analyses are therefore particularly relevant and may not have been covered in general introductory courses.

Clusters

Many occupational epidemiology studies are motivated by reports of workplace clusters of disease. A cluster investigation initially involves defining the population in the time period under study (the study base) and ascertaining whether the disease occurrence in this study base is greater than expected. However, the occurrence of a statistically significant cluster of occupational disease can merely establish a hypothesis that requires further scrutiny in another study base. For this reason, it is often argued that the investigation of disease clusters is a public health 'social service' but has little scientific value. However, several causes of chronic occupational disease were discovered through cluster investigations and such investigations will continue to be an important method for identifying new causes of chronic occupational disease (Crane et al. 1994). Cluster investigations will also be of particular interest for occupational physicians and public health officers since they will regularly encounter cluster reports in the course of their work. In particular, occupational health professionals, as well as general practitioners, can play a major role in identifying and investigating such clusters (Merletti and Comba 1992). This can be illustrated by an exercise in which a cluster report is presented and a group of course participants is asked to define the cluster (in terms of the population at risk and the outcomes which will be considered) and design an appropriate study (in the defined population and/or another population) as well as an interim press statement describing the cluster and what further action is being taken.

Incidence cohort studies

Occupational epidemiology frequently involves rare exposures, and the availability of personnel records and historical industrial hygiene information makes historical cohort studies much more feasible than in many other branches of epidemiology.

Cohort studies are usually based on personnel records, and the importance of storing such records in perpetuity should be emphasized. However, they can also be based on registries for exposures such as asbestos, and vinyl chloride. Such registries can also play a valuable role in raising awareness of workplace hazards and stimulating preventive measures (Ahlo et al. 1988; Brooke et al. 2006).

In occupational cohort studies, methods of cohort enumeration (using personnel records), exposure ascertainment (using job exposure matrices), and follow-up (using national or regional mortality and/or disease registration records) are frequently different from community-based cohort studies. Other issues include verification of the completeness of cohorts using alternative data sources and verification of vital status using a variety of data sources (e.g. death registrations, superannuation records, electoral rolls, driver's licence records). It is important to discuss cohort studies in considerable practical detail, including practical examples of individual exposure assessment, exercises involving the actual calculation of individual person-time data and its use in calculating expected mortality and/or incidence, since the use of such data is often difficult to understand without practical experience.

Incidence case-control studies

Case-control studies can then be introduced in the context of studies nested within a defined occupational cohort (Pearce 2012a). It can be readily demonstrated that a nested case-control study can be considerably more efficient than a full cohort analysis, with no loss of validity and only a minimal loss in precision (Checkoway et al. 2004). The cohort (study base) still must be defined and enumerated, and incident cases must be identified. However, it is only necessary to collect exposure history information on the cases and on a sample of controls selected from the cohort, rather than on the entire cohort. Nested case-control studies thereby combine the advantages of cohort studies for investigating rare exposures, and of case-control studies for investigating rare outcomes.

The effect measure which the odds ratio obtained from nested case-control studies will estimate depends on the manner in which controls are selected (Vandenbroucke and Pearce 2012). Although occupational case-control studies can best be introduced and discussed in the context of nested studies, there are also some occupational exposures (e.g. farming) which are relatively common

but for which it is difficult to identify and enumerate a historical cohort. In this situation, population-based case-control studies can be conducted based on a geographically defined population. Although such studies have traditionally been presented in terms of cumulative sampling, in fact most case-control studies actually involve density sampling (often with matching on a time variable such as calendar time or age) and therefore estimate the rate ratio without the need for any rare disease assumption (Vandenbroucke and Pearce 2012).

Prevalence studies

Another feature of occupational epidemiology is the relatively frequent use of cross-sectional studies of non-fatal occupational diseases such as occupational asthma (Pearce et al. 1998). For example, a prevalence study, or a prevalence case-control study, may estimate the prevalence odds ratio for asthma associated with a specific occupational exposure such as welding fumes. This approach is often appropriate and convenient but the healthy worker effect may be of particular concern, since workers with work-related symptoms may have left employment (Checkoway et al. 2004). This problem may be particularly strong for non-fatal chronic diseases such as occupational asthma (Eisen 1995). Furthermore, Krzyanowski and Kauffmann (1988) have noted that most such studies have focused on industrial groups with high levels of exposure and that these groups may be particularly affected by selection effects (Graham and Graham-Tomasi 1985).

Teaching method and format

As indicated above, the study design options in occupational epidemiology are the same as for other fields of epidemiology, the emphases given to specific study designs, and the methodological issues involved, may differ. Therefore, much of the formal content of an occupational epidemiology course may cover the same general topics as for a general epidemiology course, including the various study design options, study design issues, and detailed consideration of the different possible study designs (see 'Methodological issues'). However, the detailed content will usually differ considerably.

As noted in 'Teaching objectives', it is useful to teach these methods in the context of developing a research protocol and reporting on research findings. This can be done as an individual activity but is often more valuable as a group activity. Thus, the course participants can be divided into groups early in the course and these groups can hold regular meetings to develop a research question and design an appropriate study.

This approach can be applied to almost any course at any level, with appropriate variations in the amount of detail and methodological rigour required.

Even with an introductory course with five to six sessions, it is usually possible for the participants to divide into groups, develop a hypothesis, design a basic protocol, and present it at the final session. With a longer and more advanced course, it is often possible to develop a fully fledged protocol. This process can be assisted by

- regular presentations (interspersed with the methodological teaching) from managers, unions, and government agencies regarding their priorities for occupational health research;
- 'consultation' meetings with the course coordinator to discuss the preliminary research question and the provisional study design;
- presentation, by each group or individual, of the research question and preliminary protocol to the other course participants; and
- presentation of the proposal to an appropriate 'client' (e.g. the Occupational Safety and Health Agency, or a particular company or union) if one is available and if there is a prospect that the proposal may lead to a real research project.

The more generic methods (see 'Study design') can then be covered in more conventional teaching sessions using the standard topics outlined above but with a particular emphasis on methodological issues relevant to the protocols under development. Relatively brief exercises on specific methods (e.g. calculation of SMRs) can then be completed on an individual basis but the bulk of practical time can be devoted to protocol development. Thus, one possible course timetable could involve sessions on

- the reasons for conducting occupational epidemiology studies;
- the major global occupational health problems and research questions;
- an introduction to developing a research protocol;
- meeting with managers, unions, and interested government agencies;
- an overview of study designs;
- an overview of issues of bias;
- measuring occupational exposures;
- omics methods for measuring exposures and/or disease;
- preliminary discussion of research protocols with the course coordinator and the course participants;
- routine mortality and morbidity analyses;
- clusters;
- cohort studies;

- case-control studies;
- cross-sectional studies;
- current topics in occupational epidemiology;
- political and ethical issues; and
- presentation of research protocols.

Assessing students' achievements

Under this approach, the methodological course exercises (e.g. calculation of SMRs) form a small part of the course work and can usually be completed on an individual basis in the evening or day following the discussion of the relevant topic. The course assessment is then primarily or solely based on the research proposal. This can be developed as an individual or as a group activity (the latter is often preferable) and can be presented in oral or in written form, depending on the resources and the time available. The protocol can then be assessed with an oral or written review, as would be provided for a grant application. The review should ideally follow as closely as possible the actual format used for reviews of grant applications from the principal occupational health research funding body (e.g. the US National Institutes of Health (NIH)).

In addition, it is often useful for course participants to draw up a few simple tables of hypothetical results for the proposed study (these tables would follow the same format as would be used in a journal publication) and write a brief press statement summarizing the results. These 'findings' can also be presented and discussed (in oral and/or written form) when the final protocol is presented.

Acknowledgements

The Massey University Centre for Public Health Research is supported by a Programme Grant from the Health Research Council of New Zealand.

References

Ahlo, J., Kauppinen, T., and Sundquist, E. (1988) Use of exposure registration in the prevention of occupational cancer in Finland. *American Journal of Industrial Medicine*, **13**: 581–92.

Armstrong, B. K., White, E., and Saracci, R. (1992) *Principles of Exposure Measurement in Epidemiology*. New York: Oxford University Press.

Axelson, O. (1978) Aspects on confounding in occupational health epidemiology. *Scandinavian Journal of Work, Environment and Health*, **4**: 85–9.

Axelson, O. (1995) Possibility that solvent exposure is a risk factor for Alzheimer's disease. *American Journal of Epidemiology*, **141**: 1075–9.

Barnes, D. E. and Bero, L. A. (1998) Why review articles on the health effects of passive smoking reach different conclusions. *Journal of the American Medical Association*, **27**: 1566–70.

Boffetta, P., Andersen, A., Hansen, J., Olsen, J. H., Plato, N., Teppo, L., Westerholm, P., and Saracci, R. (1999) Cancer incidence among European man-made vitreous fiber production workers. *Scandinavian Journal of Work Environment and Health*, **2**: 222–6.

Boffetta, P., Jourenkova, N., and Gustavsson, P. (1997) Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes and Control*, **8**: 444–72.

Brooke, D., Cowley,S.,Else,D., and Leggett,S. (2006) *International Review of Surveillance and Control of Workplace Exposures*. NOHSAC Technical Report, 5. <http://www.dol.govt.nz/publications/nohsac/techreport5/index.asp>, accessed 4 November 2014.

Checkoway, H., Pearce, N., and Kriebel, D. (2004) *Research Methods in Occupational Epidemiology*. New York: Oxford University Press.

Crane, J., Lewis, S., Slater, T., Crossland, L., Robson, B., D'Souza, W., Pearce, N.,Town, G. I, Garrett, J., and Armstrong, R. (1994) The self reported prevalence of asthma symptoms amongst adult New Zealanders.[comment]. *New Zealand Medical Journal*, **10**: 417–21.

Cullinan, P. and Pearce, N. (2011) The asbestos disease epidemic: here today, here tomorrow. *Occupational and Environmental Medicine*, **67**, 98–9.

Davey Smith, G. (2001) Reflections on the limitations to epidemiology. *Journal of Clinical Epidemiology*, **54**: 325–31.

Egilman, D. (2006) Corporate corruption of science—the case of chromium(VI). *International Journal of Occupational and Environmental Health*, **1**: 169–76.

Eisen, E. A. (1995) Healthy worker effect in morbidity studies. *Medicina del Lavoro*, **86**: 125–38.

Eisen, E. A., Holcroft, C. A., Greaves, I. A.,Wegman, D. H., Woskie, S. R., and Monson, R. R. (1997) A strategy to reduce healthy worker effect in a cross-sectional study of asthma and metalworking fluids. *American Journal of Industrial Medicine*, **3**: 671–7.

Gorrell, J. M., Johnson, C. C., Rybicki, B. A., Peterson, E. L., Kortsha, G. X., Brown, G. G., and Richardson, R. J. (1997) Occupational exposures to metals as risk factors for Parkinson's disease. *Neurology*, **48**: 650–8.

Graham, S. and Graham-Tomasi, R. (1985) Achieved status as a risk factor in epidemiology. *American Journal of Epidemiology*, **122**: 553–8.

Guidotti, T. L., ed. (2011) *Global Occupational Health*. New York: Oxford University Press.

Jennen, D., Ruiz-Aracama, A., Magkoufopoulou, C., Peijnenburg, A., Lommen, A., van Delft, J., and Kleinjans,J. (2011) Integrating transcriptomics and metabonomics to unravel modes-of-action of, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in HepG2 cells. *BMC Systems Biology*, **5**: 139.

Kamburov, A., Cavill, R., Ebbels, T. M., Herwig, R., and Keun, H. C. (2011) Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics*, **2**: 2917–18.

Krzyzanowski, M. and Kauffmann, F. (1988) The relation of respiratory symptoms and ventilatory function to moderate occupational exposure in a general population. *International Journal of Epidemiology*, **17**: 391–406.

Kukull, W. A., Larson, E. B., Bowen, J. D., McCormick, W. C., Teri, L., Pfanschmidt, M. L., Thompson, J. D., O'Meara, E. S., Brenner,D. E., and van Belle, G. (1995) Solvent exposure as a risk factor for Alzheimer's disease: a case-control study. *American Journal of Epidemiology*, **14**: 1059–71.

Kumpula, L. S., Mäkelä, S. M., Mäkinen, V. P., Karjalainen, A., Liinamaa, J. M., Kaski, K., Savolainen, M. J., Hannuksela, M. L., and Ala-Korpela, M. (2010) Characterization of metabolic interrelationships and in silico phenotyping of lipoprotein particles using self-organizing maps. *Journal of Lipid Research*, **5**: 431–9.

Liou, H. H., Tsai, M. C., Chen, C. J., Jeng, J. S., Chang, Y. C., Chen, S. Y., and Chen, R. C. (1997) Environmental risk factors for Parkinson's disease: a case-control study in Taiwan. *Neurology*, **48**: 1583–8.

Mendes, R. and Costa Dias, E. (2011) 'Health protection, health promotion and disease prevention at the workplace', in **T. L. Guidotti**, ed., *Global Occupational Health*. Oxford: Oxford University Press, pp. 340–54.

Merletti, F. and Comba, P. (1992) 'Occupational epidemiology', in **J. Olsen**, **D. Trichopolous**, and **R. Saracci**, eds, *Teaching Epidemiology*. Oxford: Oxford University Press.

Michaels, D. (2005a) Doubt is their product. *Scientific American*, **29**: 96–101.

Michaels, D. (2005b) Scientific evidence and public policy. *American Journal of Public Health*, **95**: S5–7.

Monforton, C. (2006) Weight of the evidence or wait for the evidence? Protecting underground miners from diesel particulate matter. *American Journal of Public Health*, **9**: 271–6.

Pearce, N. (1992) Methodological problems of time-related variables in occupational cohort studies. *Revue d'Epidemiologie et de Sante Publique*, **40** Suppl. **1**: S43–54.

Pearce, N. (2004) The globalization of epidemiology: introductory remarks. *International Journal of Epidemiology*, **3**: 1127–31.

Pearce, N. (2007a) *Adverse Reactions: The Fenoterol Story*. Auckland: Auckland University Press.

Pearce, N. (2007b) The rise and rise of corporate epidemiology and the narrowing of epidemiology's vision. *International Journal of Epidemiology*, **36**: 713–7.

Pearce, N. (2008) Corporate influences on epidemiology. *International Journal of Epidemiology*, **3**: 46–53.

Pearce, N. (2011a) Epidemiology in a changing world: variation, causation and ubiquitous risk factors. *International Journal of Epidemiology*, **4**: 503–12.

Pearce, N. (2011b) Registration of protocols for observational research is unnecessary and would do more harm than good. *Occupational and Environmental Medicine*, **6**: 86–8.

Pearce, N. (2012a) Classification of epidemiological study designs. *International Journal of Epidemiology*, **4**: 393–7.

Pearce, N. (2012b) Global occupational health (book review). *International Journal of Epidemiology*, **41**: 896–7.

Pearce, N., Beasley, R., Burgess, C., and Crane, J. (1998) *Asthma Epidemiology: Principles and Methods*. New York: Oxford University Press.

Pearce, N., de Sanjose, S., Boffetta, P., Kogevinas, M., Saracci, R., and Savitz, D. (1995) Limitations of biomarkers of exposure in cancer epidemiology. *Epidemiology*, **6**: 190–4.

Pearce, N. and Greenland, S. (2004) 'Confounding and interaction', in **W. Ahrens**, **K. Krickeberg**, and **I. Pigeot**, eds, *Handbook of Epidemiology*. Heidelberg: Springer-Verlag, pp. 375–401.

Pearce, N. and Matos, E. (1994) Introduction. *IARC Scientific Publications*, **129**: 1–3.

Rappaport, S. M. (2011) Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology*, **2**: 5–9.

Rappaport, S. M. (2012) Discovering environmental causes of disease. *Journal of Epidemiology and Community Health*, **6**: 99–102.

Rappaport, S. M. and Smith, M. T. (2010) Environment and disease risks. *Science*, **330**: 460–1.

Rothman, K. J. and Greenland, S. (1998) *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven.

Seidler, A., Hellenbrand, W., Robra, B. P., Vieregge, P., Nischan, P., Joerg, J., Oertel, W. H., Ulm, G., and Schneider, E. (1996) Possible environmental, occupational, and other etiologic factors for Parkinson's disease: a case-control study in Germany. *Neurology*, **46**: 1275–84.

Siemiatycki, J., Wacholder, S., Dewar, R., Wald, L., Begin, D., Richardson, L., Rosenman, K., and Gerin, M. (1988) Smoking and degree of occupational exposure: are internal analyses in cohort studies likely to be confounded by smoking status? *American Journal of Industrial Medicine*, **1**: 59–69.

Slater, T., Erkinjuntti-Pekkanen, R., Fishwick, D., Bradshaw, L., Pearce, N., Cheng, S., Armstrong, H., and McLean, D. (2000) Changes in work practice after a respiratory health survey among welders in New Zealand. *New Zealand Medical Journal*, **113**: 305–8.

Steenland, K., ed. (1993) *Case Studies in Occupational Epidemiology*. New York: Oxford University Press.

Steenland, K. and Stayner, L. (1991) The importance of employment status in occupational cohort mortality studies. *Epidemiology*, **2**: 418–23.

Swerdlow, A. J. (1990) Effectiveness of primary prevention of occupational exposures on cancer risk. *IARC Scientific Publications*, **103**: 23–56.

't Mannetje, A. and Pearce, N. (2010) Occupational mortality studies: still relevant in the, 21st century. *Occupational and Environmental Medicine*, **6**: 802–3.

Vandenbroucke, J. P. and Pearce, N. (2012) Case-control studies: basic concepts. *International Journal of Epidemiology*, **4**: 1480–9.

Vlaanderen, J., Moore, L. E., Smith, M. T., Lan, Q., Zhang, L., Skibola, C. F., Rothman, N., and Vermeulen, R. (2010) Application of OMICS technologies in occupational and environmental health research; current status and projections. *Occupational and Environmental Medicine*, **6**: 136–43.

Wild, C. P. (2005) Complementing the genome with an 'exposome': the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers and Prevention*, **14**: 1847–50.

Chapter 9

Life course epidemiology

Yoav Ben-Shlomo and Diana Kuh

Introduction to life course epidemiology

Background to life course epidemiology

A life course approach to epidemiology is the study of the long-term effects on later health or disease risk of physical or social exposures during gestation, childhood, adolescence, young adulthood, or later adult life (Kuh et al. 2003). Much of the interest in life course epidemiology has centred around chronic diseases such as coronary heart disease (CHD), type II diabetes, and cancer (Trichopoulos 1990; Adami et al. 1995; Kuh and Ben-Shlomo 2004) but its concepts have also been adopted for mental and dental health (Nicolau et al. 2007). Some of the questions raised by a life course approach include the following: do foetal growth and development influence the risk of CHD and is this modified by adult obesity? Does the timing of puberty have a long-term programming effect on various endocrine pathways that may increase risk of cancer? Does childhood obesity have the same metabolic consequences and long-term effects as adult obesity and when is the best time to intervene?

For some, life course epidemiology is a new discipline, while those with longer memories often regard it as the re-emergence of ideas that have been around, in some guise or other, since the inception of the discipline. Epidemiology has a long history of considering early life exposures in relation to later disease (for a more detailed review, see Kuh and Davey Smith 2004). The study of risk factors for adult tuberculosis highlighted the role of childhood infection and long latency periods. Similarly, Donald Reid (1969) suggested that 'the bronchitic child is father to the bronchitic man'. The seminal prospective study of doctors' smoking behaviour (Doll and Peto 1976; Doll et al. 1980) also implicitly recognized the role of smoking over many years, and usually initiated in early adulthood, for many chronic diseases. Even more subtle interactive effects of childhood and adult weight on later disease risk were already being considered by Sidney Abraham and his colleagues in 1971 when they used a prospective design to show that the highest rates of hypertensive cardiovascular

disease were seen in overweight adults who were below average weight as children (cited in Kuh and Davey Smith 2004).

We feel that there have been three major areas of development in recent years that have made life course epidemiology emerge as a more coherent intellectual strand, worthy of both research and educational endeavour. The primary driver has been the ability to empirically test life course hypotheses. This has been made possible by the elegant use of historical cohorts and record linkage studies, and more recently from prospective cohort studies as well. There are now a variety of such studies that have generated consistent empirical evidence on the association between early life exposures and later disease (Kuh and Ben-Shlomo 2004). Second, there has been the development of new statistical methods and epidemiological thinking in relation to causal models that can be usefully applied to etiological questions framed within a life course paradigm (De Stavola et al. 2006). Though most of these methods were initially developed for other purposes, life course datasets have acted as a stimulus for their use and as further impetus for subsequent methodological developments. Finally, there has been the explicit recognition of conceptual models of life course pathways, absent in earlier studies, which have tried to bridge across other epidemiological models (Ben-Shlomo and Kuh 2002; Kuh et al. 2003).

Life course in other disciplines

While epidemiologists are rather new converts to life course thinking, there is a tradition in sociology, demography, and developmental psychology to consider the long-term influences of events operating across life at both an individual and a macro level. Key intellectual thinkers in these areas include Glen Elder and Paul Baltes (see Settersten 2002 for a review of life course approaches to sociology and psychology). Epidemiologists using a life course approach have, to varying degrees, tried to integrate macro, individual, and micro level variables, whereby models may include ecological measures of childhood area characteristics, through to individual micro level measures of physiology and biochemistry. It is of value for epidemiology students to have some cognizance of this broader area of life course research, even if the outcomes of interest are rather different. This promotes truly interdisciplinary research but is also useful because some of the concepts from other disciplines may be applied to epidemiological problems. However, students should be aware that confusion can occur when epidemiologists and other social scientists use the same technical terms but with rather different meanings. We have previously produced a glossary of life course terms to help avoid some of this confusion as well as to establish a common set of terms (Kuh et al. 2003).

In the following sections, we will discuss educational aspects of life course epidemiology from both an undergraduate and postgraduate perspective.

Undergraduates

Teaching objectives

In this section, we have focused on medical undergraduates, as this is our area of educational experience. Clearly undergraduates reading sociology, psychology, or demography may also consider life course approaches, although they will probably consider this within the context of their own disciplinary tradition. Some medical students do an additional year for an intercalated bachelor of science (BSc). This allows them dedicated time to study an area in far greater depth. Such courses would allow students to study life course epidemiology in far greater depth, similar to a postgraduate course.

Medical schools in the UK have to ensure that their medical programmes are consistent with the guidance of the General Medical Council's 'Tomorrow's Doctors' (General Medical Council 2003). This document highlights the range of knowledge, skills, and attitudes required of newly qualified medical practitioners. Not surprisingly, there is no specific reference to a life course approach. The document contains a section entitled 'the scientific basis of practice', which states the following:

> They [medical students] must know about and understand normal and abnormal structure and function, including the natural history of human diseases, the body's defence mechanisms, disease presentation and responses to illness. This will include an understanding of the genetic, social and environmental factors that determine disease and the response to treatment. (General Medical Council 2003)

This rather broad statement is so all-encompassing that it presents an open door to any medical educator who wishes to include life course epidemiology in their curriculum. Two aspects are particularly appropriate: (1) understanding the normal trajectories of physiological systems across the life course (see Fig. 9.1), and (2) teaching students how early life environmental influences impact on later life disease risk.

Teaching content

The medical degree syllabus is already so overloaded that one must seriously consider whether adding any more, without removing content, is justifiable. A more sensible approach is simply to build on or adapt the existing syllabus. For example, many medical schools will have a teaching session on inequalities in health. This is usually descriptive, although it may also explore possible explanations and public health policies that may reduce such inequalities. A simple

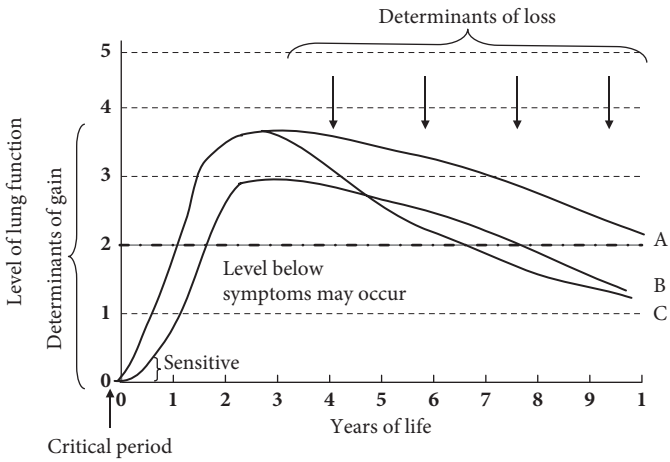


Fig. 9.1 Schematic representation of lung function over the life course. (A) represents an individual who never develops respiratory symptoms. (B) represents someone with suboptimal development of lung function but normal age-related decline. (C) represents someone with symptoms due to more rapid adult decline but normal development.

Adapted from Y. Ben-Shlomo and D. Kuh, A lifecourse approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives, *International Journal of Epidemiology*, Volume 3, Issue 2, pp. 285–93, Copyright © International Epidemiological Association 2002, by permission of Oxford University Press

approach to raising awareness of life course concepts would be to highlight the cumulative effects of lifetime socio-economic status on health, and the effects of social mobility in altering risk factors and disease risk. A life course approach could also be integrated in any session that examines migrant studies, as place of birth and place of residence in adulthood can be used to examine for critical/sensitive period effects (Strachan et al. 2007).

Another potential area of teaching would be to get students to understand how physiological function changes with age. We have previously used the example of lung function to understand how there is a developmental phase, then a plateau phase, and finally a decline phase within individuals. The trajectory by which individuals reach their clinical endpoint may differ depending on whether they experience exposure affecting the development phase or adult decline phase (subject B or C in Fig. 9.1). This life course perspective is generally absent in other disciplines. In clinical settings, doctors will focus on individuals who reach some threshold associated with clinical symptoms, for example, breathlessness. Embryologists and developmental biologists tend to focus on very early life (although there have been some exceptions (see Kuh and Davey

Smith 2004)). Respiratory physiologists help students understand the principles behind cross-sectional measures of lung function (e.g. FEV_1 and FVC) but may ignore longitudinal changes. Paediatricians will teach students about the natural history of conditions such as childhood asthma but they tend not to go beyond early adulthood. There is often, therefore, no real appreciation of the changing nature of such measures. For example, while all students are taught how to diagnose hypertension and treat it, few are aware of how the nature of blood pressure changes with age so that essential hypertension, the major cause, in midlife is overtaken by isolated systolic hypertension in older age.

Teaching methods and format

Most undergraduate medical education is done through large-group lectures in the early years and small-group bedside teaching in the later years. At the University of Bristol, we dropped all large-group lectures in favour of small-group (twelve to fifteen students) tutorials supported by computer assisted e-learning materials. This encourages a more active engagement of students through interactive sessions that allow students to think through concepts. Of course, such an approach is far more staff costly and small departments may not be able to find enough teaching staff.

A particularly useful format is a mini project (known in the UK as a Student Selected Component (SSC)), where students are given time to work on their own but under the supervision of an individual teacher. These tend to be library-based projects developing the student's critical appraisal skills, although in some cases students have undertaken their own secondary data analysis of an existing dataset. The main limitation here is that most students are unfamiliar with a specific statistical package and only a few highly motivated students are likely to choose life course topics.

Assessing students' achievements

The standard method of assessment is through written examinations. These tend to contain multiple-choice questions (MCQs), short-note, or essay questions. At Bristol, we also use extended MCQs (E-MCQs), where students must identify the correct response from a list of twenty-six possible answers rather than simply choosing a TRUE/FALSE response. This is more challenging as the student is prompted less, making guessing harder. However, life course concepts are not ideally suited to this format. In our main exam paper we have two sections: the first is an abridged version of a paper, which students must critically appraise. We currently choose more clinical scenarios but life course papers could be used. The second section involves a 'design, a study to test a hypothesis' question, which is more suited to appraising a student's understanding of life

course issues. For example, they should appreciate that, for most early life hypotheses (but not all; e.g. breastfeeding), it is not ethical or practical to do a randomized controlled trial to look at long-term outcomes. A prospective study, if already in existence, would be an excellent design but in its absence could be substituted by a historical cohort study. Students would need to consider the problems of identifying such a cohort, the limited data on potential confounders and issues around data linkage or individual follow-up. The SSC projects use standardized marking schemes, which can assess originality, creativity, and effort.

Conclusions

The medical undergraduate curriculum offers opportunities to incorporate life course concepts but there are many competing demands for student time. Key concepts of this approach could easily be integrated within existing courses on epidemiology and public health, though ideally would also cut across developmental sciences, physiology, and specific clinical topics, as appropriate (e.g. bone mass across the life course in relation to osteoporosis could be covered by rheumatology whereas early life cognitive function and risk of dementia could be covered by geriatricians or neurologists). While epidemiologists and public health teachers may be supportive advocates, it is unlikely that basic scientists and clinicians would be enthusiasts. As such, we would recommend a subtle, 'light touch' approach when incorporating life course epidemiology.

Postgraduates

We are assuming that postgraduates would be students undertaking a Masters in Epidemiology or Public Health. However, the concepts covered could also be included as either modules or a specific session within other postgraduate courses such as medical statistics or medical sociology. In the former case there would clearly be a greater emphasis on analytical issues and different statistical methods. Many of the concepts in life course epidemiology are built around existing epidemiological concepts (e.g. interaction) and assume knowledge of standard epidemiological designs. There is also a strong conceptual link to the ideas behind 'causal modelling' (Hernán et al. 2002; see also chapter 3), which is usually taught on more advanced courses.

Teaching objectives

The main teaching objectives that need to be conveyed are around different hypothetical models that may operate for exposures acting at different points across the life course (see Fig. 9.2). These diagrams illustrate the ideas behind

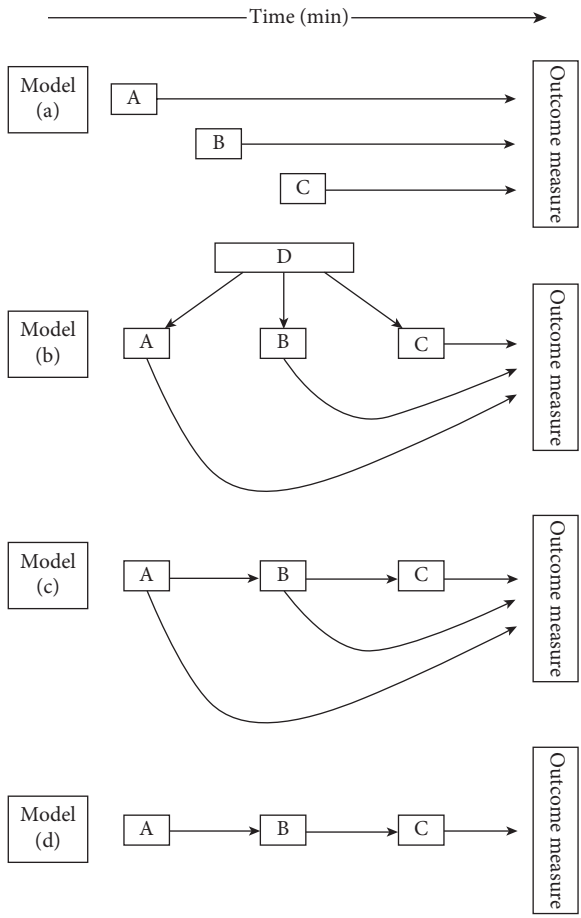


Fig. 9.2 Life course causal models.

Reproduced with permission from D. Kuh et al., Life course epidemiology, *Journal of Epidemiology and Community Health*, Volume 57, pp. 778–83, Copyright © 2003 BMJ Publishing Group Ltd.

our classification of life course models (Box 9.1). These models are necessarily an oversimplification of a complex reality and are not mutually exclusive. Students will need to familiarize themselves with some of the basic terms used to understand these models and test hypotheses. The key terms (in alphabetical order) are (1) accumulation of risk (models (a) and (b) in Fig. 9.2); (2) birth cohort effects; (3) chain of risk (models (c) and (d) in Fig. 9.2); (4) context; (5) critical period; (6) embodiment; (7) induction and latency period; (8) life cycle, life span, and life course; (9) mediating and modifying factors; (10) plasticity; (11) resilience; (12) sensitive period; (13) susceptibility; (14) time; (15) trajectory,

Box 9.1 Hypothetical life course models

Critical/Sensitive-period models
- With or without effects of late life risk factors
- With effects of later life effect modifiers

Accumulation or risk models
- Independent and uncorrelated insults
- Correlated insults
  - Risk clustering models
  - Chains of risk models
    - Additive models
    - Trigger models

Reproduced from Y. Ben-Shlomo and D. Kuh, A lifecourse approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives, *International Journal of Epidemiology*, Volume 3, Issue 2, pp. 285–93, Copyright © International Epidemiological Association 2002, by permission of Oxford University Press.

transition, and turning point; and (16) vulnerability. For explanations of these concepts, see Box 9.2 and the article by Kuh et al. (2003).

A distinction needs to be made between critical/sensitive-period and accumulation models. When interpreting evidence, these models are often confused. For example, some researchers claim to have evidence of early life programming, that is, a critical period model when they find that an early life variable remains a predictor in a model that has adjusted for later life variables (O'Connor et al. 2003). However, as the later life variables may be further measures of the same exposure, this is misleading and the results may be more compatible with an accumulation model. The key issue relates to the timing and duration of exposure. Epidemiologists naturally test for duration in their models by looking for 'dose-response effects' (a familiar concept to epidemiological students). However, the concept of dividing exposure time into specific slices of 'exposures years of risk' which may or may not have different biological impacts is less familiar. If an exposure produces the same effect on risk outside a postulated 'biological window' as it does inside that window, then the period is neither critical nor sensitive. Similarly, exposure that starts in early life and is maintained will have a greater effect simply due to its longer period of duration.

Box 9.2 Glossary of terms

Accumulation of risk—the notion that life course exposures or insults gradually accumulate through episodes of illness and injury, adverse environmental conditions, and health damaging behaviours

Birth cohort effects—An environmental change (such as an improvement or deterioration of living standards) that affects a birth cohort of individuals as indexed by their year of birth and that may show up several decades later as birth cohort differences in mortality or morbidity

Chain of risk—A chain of risk model refers to a sequence of linked exposures that raise disease risk because one bad experience or exposure tends to lead to another and then another

Context—Context refers to the location of an individual by time and place. Place refers to both geographical location and to group membership in terms of family, friends, or age, and on the basis of class, ethnicity, residence, and gender that arise out of the social and economic structure of society

Critical and sensitive periods—a critical period as a limited time window in which an exposure can have adverse or protective effects on development and subsequent disease outcome. Outside this developmental window there is no excess disease risk associated with exposure. A sensitive period is a time period when an exposure has a stronger effect on development and subsequent disease risk than it would at other times. Outside the time period any excess risk will be weaker

Embodiment—Embodiment describes how extrinsic factors experienced at different life stages are inscribed into an individual's body functions or structures

Induction and latency period—Induction period is defined as the time between exposure and initiation of the disease process

Latency period refers to the period between disease initiation and detection, and is a characteristic of the disease and/or the healthcare system.

Life cycle and life span—The concept of a life cycle has generally been used to describe a series of distinct, bounded life stages which are socially or biologically determined. In contrast the concept of the life span used in psychology assumes that development and ageing form a continuous process from birth to death.

Mediating and modifying factors—A risk or protective factor mediates the association between exposure and disease when it chronologically follows the exposure and is conceptualized as lying, at least partly, on the causal

Box 9.2 Glossary of terms *(continued)*

pathway A risk or protective factor modifies the association between an exposure and disease when the causal effect of the exposure of interest differs across levels of the modifying factor

Plasticity—Plasticity is the potential for change in intrinsic characteristics in response to environmental stimuli

Resilience—Resilience is a dynamic process of positive adaptation in the face of adversity

Susceptibility and vulnerability—Within epidemiology susceptibility has been defined in terms of Rothman's "'causal pie model'", which describes susceptibility as a process occurring over time that may eventually lead to disease via the completion of the last piece of the causal pie

Trajectory—A trajectory provides a long-term view of one dimension of an individual's life over time. These may be social, psychological states or physiological states.

Transition and turning point Transitions are short-term changes that are embedded in trajectories. A marked change of direction in a trajectory is referred to as a turning point

One example that we have used is a breast cancer case-control study which noted that smoking initiated within five years of menarche was associated with an increased risk of pre-menopausal breast cancer (odds ratio 1.77, 95% confidence interval 1.22, 2.57, $P = 0.002$) as compared to post-menopausal breast cancer (odds ratio 0.81, 95% confidence interval 0.60, 1.09, $P = 0.16$; Band et al. 2002). The authors postulated that polycyclic aromatic hydrocarbons found in tobacco smoke have greater carcinogenic effects when metabolized by breast tissue during a period of rapid cellular proliferation around puberty. This interesting observation requires further replication, especially as the authors do not mention whether adjustment for smoking years attenuated the association but, if robust, is a nice example of how the same exposure may have different effects depending on the timing rather than the duration of exposure.

We have recently further refined the above models to incorporate ideas around ageing phenotypes. While the models specified above may enhance our etiological understanding, they do not accurately represent the dynamic interplay of developmental, risk factor (behavioural or environmental) and

ageing-related trajectories. Students should be encouraged to consider why individuals age so differently even when they have similar life circumstances and genetics; for example, siblings from the same family (Plomin and Daniels 1987). Our models need to be able to take into account the large degree of variation in resilience to environmental age-related challenges. We have identified two at least sources of resilience to ageing post maturity (Kuh et al. 2014). The first is intrinsic 'compensatory reserve'; this may reflect structural factors, for example, size of kidney, functional capacity, that is, the efficiency of the kidney at filtering blood, and/or recuperative function, that is, the ability of body systems to compensate physiologically or repair damage when faced with acute or chronic low level challenges, for example, the effect of high blood pressure on kidney function. We envisage compensatory reserve as an intrinsic biological phenomenon which preserves function but deteriorates with age, probably in a non-linear fashion and possibly in parallel across multiple domains (e.g. vascular, neurological, immune, homeostatic systems) in those who will become frail. For example, the degree of neurological damage after a stroke will not only depend on the location and size of brain damage but also on the pre-morbid structure and function and the ability to open up collateral circulation and brain plasticity so that neighbouring areas can to some degree compensate for neuronal loss.

The second aspect of resilience in ageing is the individual's extrinsic responses or adaptations when faced with age-related declines, by altering behaviour (e.g. dietary intake) or the environment (e.g. move to different neighbourhood) to modify the effect on, or slow the rate of, functional decline. So for example, the 'use it or lose it' hypothesis argues that continued cognitive stimulation will maintain synaptic connectivity despite neurodegenerative changes and hence slow down decline or maintain cognitive function due to enhanced cognitive reserve. Students should consider how an epidemiological study would attempt to measure these dimensions and consider the relevance of expertise in other disciplines such as psychology, sociology, etc.

One of the features of the life course perspective is the ability to detect early markers of decline, because this offers an opportunity to intervene and potentially delay the onset of disease by modifying the rate of change through preventive strategies. Another is the ability to capture environmental characteristics before impairments emerge, facilitating a better understanding of the unfolding interaction between individuals and their environments.

A life course perspective requires repeated measures that capture both the internal and external environment, reflecting the growing interest in the 'exposome' to complement the genome (Wild 2012). Students should be excited by the prospects offered by new technologies which can help us better characterize

(a) the external environment, for example, environmental pollutants using sensors, behaviour, for example, physical activity using accelerometers and experiences (e.g. real-time logs of emotional stressors); and (b) the internal environment using, for example, biomarkers of effect as intermediate phenotypes and biomarkers of mechanisms to capture causal processes. Methodologies such as proteomics and epigenomics will complement the genomic data and give us a better handle on exposures that are often difficult to capture such as toxic exposures.

Teaching content

Any educator has to make a choice as to whether they wish to integrate life course epidemiology within other aspects of the syllabus or make it a module in its own right. This will clearly depend on the interests of staff and the degree to which life course epidemiology is seen as a teaching priority. There may be an optional life course module, covering between four and eight sessions, some of which are lecturer led while others are student led or use peer or self-directed learning. A possible syllabus for such a module would cover all or some of the following:

1. The history of life course epidemiology and its relationship to other disciplinary approaches—the session would describe how epidemiological thinking has evolved (Susser 1985) and how the focus on risk factors at different points in the life course has changed over time; it would also describe parallel developments in other disciplines such as sociology and psychology.
2. Life course models and concepts—this session would cover the main concepts and models and ensure students understand the key terminology.
3. Life course designs—how one can test life course hypotheses using a range of study designs from conventional (historical cohort studies) to less conventional approaches (twins or sibs discordant on exposure of interest). Some courses could also link this to advances in genetic epidemiology and how the use of functional genetic variants can act as an unconfounded proxy life course exposure (see Davey Smith and Ebrahim 2003 for a discussion of Mendelian randomization). This session is best taken as an exercise around a specific hypothesis introduced by the lecturer; the students would then explore different design strategies, weighing up their respective strengths and weaknesses (see 'Teaching methods and format' for a detailed case study).
4. How life course relates to other epidemiological models—this session would consider other epidemiological modes of thinking, particularly the interrelationship between life course epidemiology and causal models and the need

to consider clearly the temporal relationship between variables. In addition, life course epidemiology has been applied to social epidemiology and has sometimes been considered as an alternative to eco-social models (Susser and Susser 1996a, b; Krieger 2001). It is true that most life course research until now has tended to ignore macro level factors and focus mainly on individuals but this is usually due to a limitation of the available data rather than of the concepts. In fact, multilevel models that have examined the role of contextual factors tend only to consider this at across-sectional levels and ignore that areas also have their own life courses and change over time. It is therefore perfectly possible to integrate both approaches (see Ben-Shlomo and Kuh 2002 for a discussion of this topic).

5. Statistical issues in life course epidemiology—this could span far more than one session, depending on prior statistical knowledge. Some of the key issues are generic to any longitudinal study, for example, handling missing data; while other issues relate to understanding pathways, for example, structural equation models. Some papers have embraced the challenges of using more sophisticated methods to model repeat exposure data, for example, changes in anthropometric data due to growth (De Stavola et al. 2006), or using repeat measures of phenotypes to better capture the heterogeneity of any outcome and hence possible differences in risk factor associations (Colman et al. 2007). Methodological work has also considered how one can empirically differentiate between different life course models (Mishra et al. 2009). The methodological issues around life course epidemiology are sufficiently broad to allow a course in its own right if one desired (see Pickles et al. 2007). Pickles and De Stavola (2007) highlight the distinctions between variable-based, individual-based, and multivariate profile analysis, where the outcome is a profile of measures over an extended time period. They also discuss causal models and directed acyclic graphs (DAGs) and cover more complex statistical methods, marginal structural models, propensity scores, G-estimation, instrumental variable analysis, and structural equation models. Some of these topics will be beyond the remit of some epidemiological courses but would fit nicely into a medical statistics master's programme.
6. Specific case studies or topics of controversy—although life course research is still very much in its infancy, there are several topics that have generated much debate. Such topics could be used either as a for/against debate led by the students or as an exercise where students need to read the key papers beforehand and which would form the basis for a discussion group. Some potential examples are the relative importance of foetal growth as measured by birth weight for adult blood pressure as compared to adult dietary salt

consumption. Another topic of interest is the interpretation of the inverse association between birth weight and adult outcomes after adjustment for adult obesity (see Lucas et al. 1999; Cole et al. 2001; Osmond et al. 2001; Cole 2005; Tu et al. 2005; Weinberg 2005).

7. Understanding social inequalities in health from a life course perspective—this session examines how the patterning of disease and intermediary risk factors may be related to childhood and adult socio-economic status (Davey Smith and Lynch 2004).
8. How does life course epidemiology influence public health policies—this is a more speculative and challenging session that allows students to think about the practical implications of life course studies. For example, it has been argued that small babies are more likely to become obese children, obese adults, develop diabetes, and die of heart disease. Should one aim to intervene in childhood or wait until adulthood? Is there evidence that interventions in earlier life are more effective or will persist into adulthood? How does this compare to a population-based approach that aims to shift the whole adiposity distribution? This assumes students are aware of the approach to individual- versus population-based strategies (Rose 1985), that is, the distinction between attributable rate/risk and population attributable rate/risk.

Teaching methods and format

Most postgraduate courses will involve standard lectures as well as small-group tutorials. Because postgraduate students are usually highly motivated, there is scope for distance-learning, self-directed learning, or student-led sessions. A usual approach involves a thirty- to sixty-minute lecture, followed by a practical session that involves students working together with some teacher supervision or a debrief. Practicals can involve reading key papers and then discussing them or setting up students to present different perspectives or constructing a formal debate.

E-learning packages and websites are also useful tools, though at the moment we are not aware of any specific websites for life course epidemiology. Another method is to record lectures or seminars using software which combines audio or video with PowerPoint-type presentations. These can then be archived on the university server or virtual-learning environment for access at any time.

It is helpful to elucidate the theoretical and practical issues around undertaking life course epidemiology with a meaningful case study that would be pertinent to teachers and students. Helpful examples from a variety of chronic diseases can be found in various edited books (e.g. Kuh and Hardy 2003; Kuh and Ben-Shlomo 2004). These cover a wide range of conditions: cardiovascular

disease, diabetes, obesity, blood pressure, respiratory and allergic disorders, cancer, neuropsychiatric outcomes, reproductive health, musculoskeletal disorders, and ageing. For example, a case study that we have used involves understanding the association between insulin-like growth factor I (IGF-I), growth and development, and the future risk of specific cancers (Ahlgren et al. 2004). Even longevity has been studied by both demographers and gerontologists with particular interest in how early life factors such as season of birth (Doblhammer and Vaupel 2001) and cumulative adverse life course events (Gavrilov and Gavrilova 2003) can have a detrimental effect.

Assessing students' achievements

Most postgraduate courses will use a combination of in-course assessment as well as exam papers. As demonstrated by the case study above, it is possible to use a specific example to generate a mini project that students could work on in groups or individually and then hand in a written piece of work and/or give an oral presentation for in-course assessment. Areas that lend themselves particularly well to this are designing a study or writing a grant proposal. The latter adds the extra complexity of planning a protocol and budget. Where available, providing students with a real or simulated dataset allows them to tackle the concepts from a hands-on perspective and forces them to operationalize concepts into clear analytical strategies. Growth data are particularly useful as they are readily available and allow students to experiment with different analytical methods, from life course plots (Cole 2007) to more complex procedures (Pickles and De Stavola 2007).

Conclusions

Life course epidemiology, as a distinct area of research, is relatively new and provides methodological challenges for both epidemiologists and statisticians. It builds on conventional epidemiological theory and basic concepts such as confounding, interaction, induction, and latency period, and these concepts must be covered first. We would suggest that life course studies provide useful teaching material that challenges students to think beyond black-box epidemiology. It also promotes consideration of the web of disease causation. More importantly, it places this web in a time-dependent narrative that uniquely unfolds for individuals embedded within different social and environmental milieus. Our experience is that students appreciate and are motivated by the real and complex problems that life course epidemiology addresses and gain a deeper understanding of the reasons for variation in the health of individuals and populations.

Acknowledgements

We would like to thank Rachael Cooper for helpful comments on an earlier draft, and the many students that we have taught over the years and who have always stimulated and challenged us.

Useful websites

The following links are to some of the major birth cohorts that could provide valuable information about life course influences on health:

- MRC Unit for Lifelong Health and Ageing & MRC National Survey of Health and Development (<http://www.nshd.mrc.ac.uk/>). This is the first UK national birth cohort that has followed up individuals since 1946.
- The Centre for Longitudinal Studies (<http://www.cls.ioe.ac.uk/>) houses data on the 1958 National Child Development Study, the 1970 British Cohort Study, and the Millennium Cohort Study.
- Avon Longitudinal Study of Parents and Children (<http://www.alspac.bris.ac.uk/>) is a contemporary birth cohort that has very detailed phenotypic data and a very rich source of biological samples, including cell lines on parents and offspring.
- The Danish National Birth Cohort (<http://www.ssi.dk/sw9314.asp>), which was set up between 1997 and 2002, is one of the largest birth cohorts in the world.
- The *International Journal of Epidemiology* (<http://ije.oxfordjournals.org/>) regularly publishes cohort profiles. Some of these will be directly relevant to life course studies.
- The following website is about improving the quality of health research and useful collates reporting guidelines such as CONSORT, STROBE, STARD, QUORUM, and MOOSE (<http://www.equator-network.org/index.aspx?o=1032>), which is helpful for students.

References

Adami, H. O., Persson, I., Ekbom, A., Wolk, A., Pontén, J., and Trichopoulos, D. (1995) The aetiology and pathogenesis of human breast cancer. *Mutation Research*, **333**: 29–35.

Ahlgren, M., Melbye, M., Wohlfahrt, J., and Sørensen, T. I. (2004) Growth patterns and the risk of breast cancer in women. *New England Journal of Medicine*, **351**: 1619–26.

Band, P. R., Le, N. D., Fang, R., and Deschamps, M. (2002) Carcinogenic and endocrine disrupting effects of cigarette smoke and risk of breast cancer. *Lancet*, **360**: 1044–9.

Ben-Shlomo, Y. and Kuh, D. (2002) A lifecourse approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, **31**: 285–93.

Cole, T. (2005) Re: why evidence for the fetal origins of adult disease might be a statistical artifact: the 'reversal paradox' for the relation between birth weight and blood pressure in later life. *American Journal of Epidemiology*, **162**: 389–95.

Cole, T. (2007) 'The life course plot in life course analysis', in A. Pickles, B. Maughan, and M. Wadsworth, eds, *Epidemiological Methods in Life Course Research* (1st edn). Oxford: Oxford University Press, pp. 137–56.

Cole, T. J., Fewtrell, M., and Lucas, A. (2001) Early growth and coronary heart disease in later life: analysis was flawed. *British Medical Journal*, **323**: 572.

Colman, I., Ploubidis, G. B., Wadsworth, M. E., Jones, P. B., and Croudace, T. J. (2007) A longitudinal typology of symptoms of depression and anxiety over the life course. *Biological Psychiatry*, **62**: 1265–71.

Davey Smith, G. and Ebrahim, S. (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, **32**: 1–22.

Davey Smith, G. and Lynch, J. (2004) 'Life course approaches to socioeconomic differentials in health', in D. Kuh **and** Y. Ben-Shlomo, eds, *A Life Course Approach to Chronic Disease Epidemiology* (2nd edn). Oxford: Oxford University Press, pp. 77–115.

De Stavola, B. L., Nitsch, D., dos Santos Silva, I., McCormack, V., Hardy, R., Mann, V., Cole, T. J., Morton, S., and Leon, D. A. (2006) Statistical issues in life course epidemiology. *American Journal of Epidemiology*, **163**: 84–96.

Doblhammer, G. and Vaupel, J. W. (2001) Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences USA*, **98**: 2934–9.

Doll, R., Gray, R., Hafner, B., and Peto, R. (1980) Mortality in relation to smoking: 22 years' observations on female British doctors. *British Medical Journal*, **280**: 967–71.

Doll, R. and Peto, R. (1976) Mortality in relation to smoking: 20 years' observations on male British doctors. *British Medical Journal*, **2**: 1525–36.

Gavrilov, L. A. and Gavrilov, N. S. (2003) The quest for a general theory of aging and longevity. *Science of Aging Knowledge Environment*, **28**: RE5.

General Medical Council. (2003) *Tomorrow's Doctors* (2nd edn). London: General Medical Council.

Hernán, M. A., Hernández-Díaz, S., Werler, M. M., and Mitchell, A. A. (2002) Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *American Journal of Epidemiology*, **155**: 176–84.

Krieger, N. (2001) Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology*, **30**: 668–77.

Kuh, D. and Ben-Shlomo, Y. (2004) *A Life Course Approach to Chronic Disease Epidemiology* (2nd edn). Oxford: Oxford University Press.

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J., and Power, C. (2003) Life course epidemiology. *Journal of Epidemiology and Community Health*, **57**: 778–83.

Kuh, D., Cooper, R., Hardy, R., Richards, M., and Ben-Shlomo, Y., eds. (2014) *A Life Course Approach to Healthy Ageing*. Oxford: Oxford University Press.

Kuh, D. and Davey Smith, G. (2004) 'The life course and adult chronic disease: an historical perspective with particular reference to coronary heart disease', in D. Kuh and Y. Ben-Shlomo, eds, *A Life Course Approach to Chronic Disease Epidemiology* (2nd edn). Oxford: Oxford University Press, pp. 15–40.

Kuh, D. and Hardy, R. (2003) *A Life Course Approach to Women's Health* (1st edn). Oxford: Oxford University Press.

Lucas, A., Fewtrell, M. S., and Cole, T. J. (1999) Fetal origins of adult disease—the hypothesis revisited. *British Medical Journal*, **319**: 245–9.

Mishra, G., Nitsch, D., Black, S., De Stavola, B., Kuh D., and Hardy, R. (2009) A structured approach to modeling the effects of binary exposure variables over the life course. *International Journal of Epidemiology*, **38**: 528–37.

Nicolau, B., Thomson, W. M., Steele, J. G., and Allison, P. J. (2007) Life-course epidemiology: concepts and theoretical models and its relevance to chronic oral conditions. *Community Dentistry and Oral Epidemiology*, **35**: 241–9.

O'Connor, T. G., Heron, J., Golding, J., Glover, V., and the ALSPAC Study Team. (2003) Maternal antenatal anxiety and behavioural/emotional problems in children: a test of a programming hypothesis. *Journal of Child Psychology and Psychiatry*, **44**: 1025–36.

Osmond, C., Barker, D. J. P., Eriksson, J. G., and Forsén, T. (2001) Early growth and coronary heart disease in later life: authors' reply. *British Medical Journal*, **323**: 572–3.

Pickles, A. and De Stavola, B. (2007) 'An overview of models and methods for life course analysis', in A. Pickles, B. Maughan, and M. Wadsworth, eds, *Epidemiological Methods in Life Course Research* (1st edn). Oxford: Oxford University Press, pp. 221–46.

Pickles, A., Maughan, B., and Wadsworth, M. (2007) *Epidemiological Methods in Life Course Research* (1st edn). Oxford: Oxford University Press.

Plomin, R. and Daniels, D. (1987) Why are children in the same family so different from one another? *Behavioral and Brain Sciences*, **10**: 1–60.

Reid, D. D. (1969) The beginnings of bronchitis. *Proceedings of the Royal Society of Medicine*, **62**: 311–6.

Rose, G. (1985) Sick individuals and sick populations. *International Journal of Epidemiology*, **14**: 32–8.

Settersten, R. A. Jr., ed. (2002) 'The study of lives: emerging propositions and controversies', in R. A. Settersten, Jr, ed., *Invitation to the Life Course: Toward New Understandings of Later Life* (1st edn). New York: Baywood Publishing Company Inc., pp. 5–64.

Strachan, D. P. et al. (2007) Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *International Journal of Epidemiology*, **36**: 522–31.

Susser, M. (1985) Epidemiology in the United States after World War II: the evolution of technique. *Epidemiologic Reviews*, **7**: 147–77.

Susser, M. and Susser, E. (1996a) Choosing a future for epidemiology: 1. Eras and paradigms. *American Journal of Public Health*, **86**: 668–73.

Susser, M. and Susser, E. (1996b) Choosing a future for epidemiology: 2. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health*, **86**: 674–7.

Trichopoulos, D. (1990) Hypothesis: does breast cancer originate in utero? *Lancet*, **335**: 939–40.

Tu, Y. K., West, R., Ellison, G. T., and Gilthorpe, M. S. (2005) Why evidence for the fetal origins of adult disease might be a statistical artifact: the 'reversal paradox' for the relation between birth weight and blood pressure in later life. *American Journal of Epidemiology*, **161**: 27–32.

Weinberg, C. R. (2005) Invited commentary: Barker meets Simpson. *American Journal of Epidemiology*, **161**: 33–5.

Wild, C. P. (2012) The exposome: from concept to utility. *International Journal of Epidemiology*, **41**: 24–32.

Chapter 10

Pharmacoepidemiology

Susan Jick

Introduction to pharmacoepidemiology

The approach to teaching pharmacoepidemiology presented in this chapter focuses primarily on techniques for conducting drug safety studies. These techniques, however, can also be applied to pharmacoeconomic and outcomes research studies, since the important principles covered in the drug safety area also form the foundation for other types of drug studies. The teaching is targeted towards students of epidemiology, as the principles involved are fairly advanced and would be difficult to cover in an undergraduate class.

Teaching objectives

Students completing this course should be able to review critically the literature in the drug safety area as well as design simple drug safety studies.

Teaching content

As with any discipline, it is important to first set out the basic principles and methods relevant to the field of study. In drug epidemiology, as in other areas of epidemiology, understanding the basic methods of clear case and exposure definition are critical to the success of a study. However, in pharmacoepidemiology, providing clear definitions is more complex than in classical epidemiology. First, drug exposures vary greatly; that is, drug use can often change within a single individual, and each drug is taken differently and therefore requires distinct exposure definitions and considerations. Thus, exposure can be difficult to define precisely. Second, the outcomes in drug safety studies are often associated in some way to the exposure and thus create opportunities for biases that are essential to anticipate and properly accommodate in the study design. Further, in this field, there are many different outcomes that can be induced by drug use; thus, one must be prepared to study many different outcomes, each of which has its own study design implications. This chapter attempts to provide the necessary material to identify the issues related to study design and to teach the methods essential to the conduct of a well-designed drug safety study.

Teaching method

I begin this course by discussing how formalized studies of drug safety were first initiated and by providing past examples of serious drug safety problems such as thalidomide and phocomelia, or isotretinoin and birth defects. There have also been important studies in this area where unintended beneficial effects of drugs have been discovered. The protective effect of aspirin on the risk for myocardial infarction is one example of such a finding (Boston Collaborative Drug Surveillance Program 1974). There are also many recent drug safety issues that many people will be aware of. I provide several references of important drug safety issues that have received attention in the medical literature (Avorn 2007; Curfman et al. 2007; Hampton 2007; Rosen 2007). Other topics include but are not limited to the issue of cyclo-oxygenase 2 (COX-2) inhibitors and the risk of myocardial infarction (the Vioxx issue), concerns about the safety of antidepressants in relation to suicide risk, and the cardiovascular safety of drugs used to treat type 2 diabetes. It is also helpful to discuss the different areas of pharmacoepidemiology such as drug utilization, outcomes research, and regulatory issues, as well as drug safety. I focus on the methods of drug safety research in this course but briefly discuss the role of these other areas.

A challenge for every course is to engage the students in the material presented in class. The exercises provided at the end of the chapter have stimulated much excitement and interest in this area of epidemiology and have inspired some very creative and thoughtful responses from the students. I encourage class participation and discussion throughout the course and challenge the students to work through the implications of using the different methodologies presented.

Below are described the key elements of drug epidemiology that need to be taught in order to begin to grasp the scope of the field and to be able to review critically the literature as well as to think about study design. With each element, there is a brief summary of the main points to be made.

Considerations for all drug safety studies

Basis for defining drug/disease relation

In order to properly design a drug epidemiology study, one must first determine the nature of the outcome and the drug to be studied. Students should be presented with the various ways of thinking about drugs and drug effects (H. Jick 1977). There are several ways to think about the exposure and the outcome in drug safety studies. Is the outcome rare or common, serious or mild? Is it a pharmacologic, idiosyncratic, or allergic disorder? Is it a functional, biochemical, or structural problem? Is it caused by short-term use, long-term use,

or both? Is the effect acute, continuous, or delayed? Is the drug newly marketed, recently marketed, or marketed for many years? All of these determinations help to frame the question at hand and consequently lead to the use of the appropriate methodology, exposure and outcome definitions, and data resources. For example, if a drug is not widely used, one will need a very large database to find enough exposed subjects whereas, if the effect is delayed, one will need to follow up study subjects in the long term in order to detect the outcomes of interest. These elements must be sorted out prior to starting a study. The teacher should go through several scenarios to help the students grasp the variety in drug and outcome types. The following are some examples that could be used to help illustrate the importance of the different drug/outcome relations.

In studies of liver disease and antibiotic use you are investigating an acute effect in what is usually short-term drug use. The exposure is common but the outcome (idiopathic liver disease) is rare. The implications of this scenario are that one needs a database with a lot of antibiotic exposure but the follow-up need not be long term. The exposure definition should be current or recent use (recent is defined as being up to forty-five days after discontinuing the drug). One must also be able to determine if the liver disease is idiopathic (i.e. not due to another proximate cause); therefore, access to the original records is required (Derby et al. 1993).

A study of oestrogens and breast cancer has different drug and disease characteristics and would require a very different design. In this scenario oestrogens can be used chronically and breast cancer is a delayed effect. Thus, one would need a database with long follow-up and continuous information on drug use. Exposure would be defined in terms of long-term use, not acute use. Information on switching of dose and hormone therapy would need to be taken into account, and information on potential confounders would be very useful. Cases would need to be defined as first-time or incident breast cancer, with no past history of cancer. Since cancer is not common, a large database would be necessary to find enough cases.

In teaching a class in pharmacoepidemiology, I carefully review the association between the drugs and diseases under investigation. Some safety studies evaluate the relation between a drug that has no clinical relation to the outcome of interest (e.g. antidepressant drugs and liver disease), while other drug–disease relations are more complicated (such as antibiotics and seizures). It is important that the disease under study not be associated with the drug under investigation. Also, the outcome must be idiopathic, that is, not due to another proximate cause. One is interested in discovering possible cases of drug-induced illness in the absence of other apparent causes. If non-drug-induced

Table 10.1 Example of how the inclusion of non-idiopathic cases in study can lead to false conclusions

Exposure	Idiopathic cases	Non-idiopathic cases*	Total cases	Number exposed	Rate of idiopathic liver disease (per 10,000)	Rate of all liver disease (per 10,000)
Antibiotic A	2	100	102	25,000	0.8	40.8
Antibiotic B	18	100	118	25,000	7.2	47.2
Relative risk for Antibiotic B vs Antibiotic A	–	–	–	–	9.0	1.2

* Note that where a non-drug cause of liver disease (non-idiopathic disease) is present, the cases are evenly distributed across the two exposures. When the cases are restricted to the idiopathic liver disease, there are few cases but they are not evenly distributed. If all cases are included in the study, since most liver disease does have another proximate cause, the true effect of Antibiotic B would be diluted by the non-drug-induced cases, and the true drug effect would be missed. Also, note that most cases of liver disease are not idiopathic; instead, cases of liver disease are most often alcohol related or due to viral hepatitis or pre-existing liver disease.

illnesses are included as cases, it is likely that a true effect will be diluted or even obscured. For example, in a study of drug-induced liver disease, one must exclude from the possible cases all persons with viral hepatitis, as this is a likely cause of liver disease and is unlikely to be drug induced. See the example presented in Table 10.1. It is helpful to provide examples of studies that are done incorrectly where a null result or minimally increased risk is found and compare it to a study that excludes non-idiopathic cases and finds a positive association between the drug and the outcome. It is also important to stress the limitations of observational studies of drug effects. Such studies cannot meaningfully evaluate an effect of a pre-existing illness, so that in a study of drug-induced seizures one must exclude all people with a prior history of epilepsy. It is rarely possible in observational studies to determine confidently if a drug causes worsening of seizures or increases the rate of seizures in someone with pre-existing disease. It would be useful at this point to have a discussion of prescribing bias where the teacher should provide (and solicit from the students) examples. One example of this type of bias is found in the study of antibiotic-associated seizures. If a subject already has a seizure disorder, does the receipt of an antibiotic worsen the condition? It is not possible to know if the antibiotic or the illness being treated by the antibiotic is causing the seizure, or if the two are unrelated. To study worsening of disease, one must conduct a clinical trial, as observational studies are inherently biased in ways that would yield uninterpretable results (H. Jick et al. 1998; Ray 1988).

Similarly, it is difficult, if not impossible, to interpret results from observational data when study drugs are used to treat diseases that are risk factors for the outcome under study using observational data. For example, the results of a study of suicide in users of antidepressants would be difficult to interpret, as people on antidepressants are depressed and would be expected to commit suicide more often than those not exposed. Further, it is likely that people taking different antidepressants have varying severity of depression, so the suicide rate would be different among the users of the various antidepressants. It is again helpful to have the students think through the logic in this example and to come up with other examples of such situations.

Control of confounding

Confounding is a concern in all epidemiologic studies and is covered in basic methodology classes. However, when I teach this course, I discuss several factors in drug epidemiology studies that must routinely be controlled, or at the very least, evaluated as potential confounders. These are age, sex, geography, and calendar time. Each of these factors is frequently correlated with the incidence of disease and almost always with drug exposure. Drug use changes from one part of the world to another and varies between males and females and among age groups. It also varies over time as new drugs become available. Rates of illness also may vary across these four factors and thus must be taken into account in study design and analysis. Providing examples of these types of variation is useful in demonstrating their potential effects on a study result.

An important technique used to control confounding is restriction. Restriction attempts to remove from a population those subjects who have an inherently higher risk of the outcome, in order to evaluate whether groups of subjects (in a case-control study, cases and controls) who are at the same or closely similar levels of risk for the outcome differ with respect to their exposure under evaluation. For this reason, it is common practice in pharmacoepidemiology to restrict studies to idiopathic cases of a disease under evaluation, in order to assess whether or not an exposure may have a causal association with the disease. Only when the risk of the outcome is comparable between cases and controls is it sensible to consider a causal relationship between an exposure that is more prevalent in the cases than in the controls and the disease of interest. Studies that evaluated the risk of fracture in users of proton pump inhibitors (PPIs) provide a helpful illustration of the difference between using restriction and controlling for confounders in the analytic phase of a study (Yang et al. 2006; Kaye et al. 2008; Gray et al. 2010). In these studies the restricted analyses found no increased risk of fracture in PPI users, while the studies that included all cases found a small increase in risk. The risks diminished when confounders

were controlled in the analyses, but did not reach the null. Review of these studies should lead to productive discussions of the different strategies for controlling confounding.

Confounding by indication

Confounding by indication is a bias that occurs when the drug of interest is selectively used or not used by those who develop the outcome of interest. This occurs when the drug is used to treat a disease that is in the causal pathway of the outcome of interest. For example, antihypertensive medications are used to treat hypertension, which is a strong risk factor for myocardial infarction. Therefore one would expect the risk of myocardial infarction to be higher in people with hypertension because of the underlying disease, not because of the exposure to the antihypertensive drug. This bias presents some of the stronger challenges in pharmacoepidemiologic studies and should be discussed thoroughly in class (Walker 1996).

Here again, working through examples will help the student grasp the nature and importance of this form of bias. Another example is found in the study of cimetidine and gastric cancer, as people with early symptoms of gastric cancer are likely to receive cimetidine to treat the symptoms. This is also known as protopathic bias. In this instance, if confounding by indication is not adequately controlled, then the researchers might incorrectly conclude that there is a positive association between the drug and the disease (Schumacher et al. 1990). Figure 10.1 provides a description of such a scenario. In this situation the problem can be solved by moving the index date, thus the date before which all exposure is assessed in this case-control study, back two years so that all drug

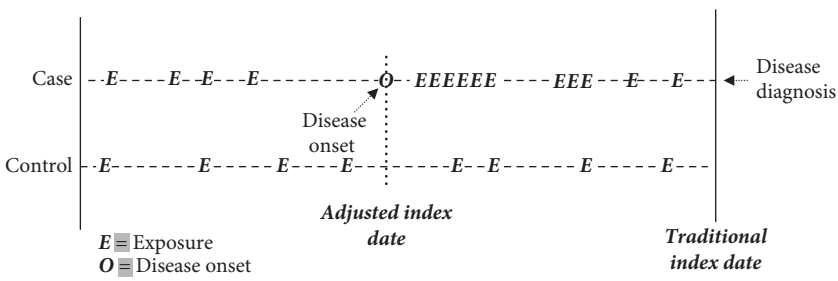


Fig. 10.1 Study design consideration to avoid confounding by indication. Note that if you assess exposure *prior* to disease onset, there is no difference in exposure between the cases and controls. However, if you look at exposure just prior to the disease diagnosis, there will be an association between the drug and the disease. This is not because the drug causes the disease but because the drug is used to treat the early symptoms of the yet-to-be diagnosed disease.

exposure would occur before onset of the disease of interest. This is also a good example of the importance of being sure that exposure precedes disease onset.

It is not always possible to control for confounding by indication, and some studies simply cannot be done or, if they are done, the results are not interpretable. When I teach this, I always describe a study that is subject to confounding by indication and then describe the proper methodology to avoid the problem (Schumacher et al. 1990). It is also helpful to work through the result that would be obtained if the proper methodology were not employed.

Exposure ascertainment

Another challenge in the area of pharmacoepidemiology is to obtain accurate and complete exposure information. The classic way to obtain the data is to ask people what drugs they have taken. It is very revealing to ask students in the class what drugs they have taken in the past three years, how long they took the drug, and what the dosage was. The point of this exercise is to demonstrate how difficult it can be to find reliable exposure information. It is then helpful to work through examples that illustrate the effects of misinformation on a study result. I like to go one step further and describe the importance of procuring complete data on exposure. It is not helpful to have information on drug use collected at one point in time ten years prior to the outcome under study. For most studies, drug exposure data must be collected continuously over time and must reflect all details of drug use, such as starting and stopping the drug, changing dose, or switching to a similar drug, up until the diagnosis of the study outcome. Most often, current use is the exposure of interest, though as discussed earlier, this is not always the case. It is important *not* to look only at ever-use compared to never-use, since ever-use would include someone who received one prescription for a drug many years prior to the index date as well as someone who was heavily exposed. These exposures should not all be considered as equivalent. It is helpful to provide examples of studies that do not have complete drug information and to discuss the implications for the study results.

A monumental improvement in drug safety studies occurred when electronic drug data became available. Detailed data on drug use are now available in many data resources including drug name, route of administration, strength, quantity prescribed, and dates of dispensing or prescribing. While these data are a major advance in this area, not all studies have access to these prospectively recorded drug details. In the absence of these drug details it is important to recognize the limitations of drug exposure data. An example can be seen in a study by Tamimi and colleagues (2006). In this study, in some of the questionnaires women were asked specifically about prior hormone use, including testosterone use, while in others testosterone was not mentioned but women were

asked to record any 'other' hormones taken (in addition to oestrogens), and the use of oestrogen plus testosterone (E + T) was not distinguished from the use of oestrogen plus testosterone plus progesterone (E + T + P). Due to the suboptimal recording of drug exposure data, the reported small increase in risk for E + T use could, in fact, reflect an excess in risk generated by progesterone. This assumption is supported by the fact that the researchers found that the effect of E + T was higher in women with a natural menopause (who would take E + T + P to prevent an increased risk of endometrial cancer) compared to women with surgical menopause (who would take E + T with no P). The results of a study of the same topic (S. Jick et al. 2009) also support this hypothesis; that is, the results indicate that E + T + P is associated with an increased risk of breast cancer while E + T alone is not. This example illustrates the importance of careful exposure ascertainment and definitions.

Considerations for cohort studies

Timing is another consideration in ascertaining exposure. It is critical that the exposure precede the outcome and, while this sounds obvious, the correct timing is not always applied. This is most often a problem when studying chronic disease when the timing of disease onset is not known. Researchers must allow for the lag between disease onset and disease diagnosis. Looking at exposure duration is important in these situations, as a duration effect should be seen if the relevant exposure window has been correctly defined. If there is an effect of short duration of use on a chronic disease but no effect of long-term use, one should consider the possibility that the association is due to some bias rather than to a causal association between the drug and the outcome. For example, in a study of non-steroidal anti-inflammatory drugs (NSAIDs) in relation to rheumatoid arthritis (RA), one would likely find an increased risk of short-term NSAID use, as people treat the symptoms of RA before they are diagnosed with the illness.

Defining the cohorts

Selecting an exposed cohort and an appropriate comparison group is crucial in the design of a study. It is important for the students in the class to think through the possible ways of defining the study cohorts. It is not always best to compare users of a drug to non-users. There are times when another drug exposure provides a more appropriate comparison. Subjects in the comparison cohort should be as similar to the exposed group as possible, except that they have not received the study drug. For example, in studying the effect of an antibiotic drug, it would be useful to compare the outcome to that of another antibiotic

used to treat the same conditions. In this way, one controls for the presence of underlying disease (the indication for drug use), which could be associated with the outcome under study. Sometimes a non-exposed cohort is the only possible or reasonable choice for a comparison group (H. Jick et al. 1998). When teaching this class, I always discuss the problems related to exposure definition that arise from changing and discontinuing drug use. Students should think through the design implications of changing exposure status, one of the challenges of pharmacoepidemiology.

Taking the students through a study and discussing the implications of these issues is very helpful in thinking through proper exposure definition. One example that illustrates this point is a study comparing different drugs for the same indication. If one is conducting a study of antidepressants in relation to breast cancer, and a study subject has been exposed to more than one antidepressant in the time before the index date (the date of diagnosis in the case), then how does one classify this subject? Since there can be many different combinations of exposure to the various study drugs, the simple approach is to restrict the study to users of only one of the study drug prior to the index date.

It is important to remind students that people receive drugs to treat medical problems. This means that, by definition, exposed study subjects will all have some medical condition (except in rare circumstances such as women using oral contraceptives). Yet, to be eligible to enter any of the study cohorts, a subject must not have had the *study* outcome prior to entry into the study. In drug safety studies, one should always look at *newly diagnosed disease.* In so doing, one is, by definition, excluding anyone who had the disease prior to receiving the study drug. For the same reason, the investigator must exclude people who have predisposing conditions that are present before entry into the study, as these conditions are the more likely cause of any subsequent development of the outcome under study and may influence exposure to the study drug (Garcia Rodriguez and Jick 1994; S. Jick 1998). This is discussed further in the section 'Considerations in case-control studies'.

Defining the exposure window

The appropriate windows of exposure must be determined for each new study separately, as it is dependent on the outcome under study and the drug under investigation. In this class, I like to go through different scenarios to be sure that the students understand how the exposure window relates to the outcomes. For acute outcomes such as liver disease, renal disease, serious skin diseases, etc., a short window such as forty-five or sixty days following discontinuation of therapy is appropriate (Derby et al. 1993). However, if one is studying the association between a drug and a delayed outcome such as cancer or cataracts, then one

might need to look at all time after exposure, taking into account duration of use, recency of use, cumulative dose, etc. (H. Jick et al. 1979; S. Jick et al. 2001).

Expressing exposure and accounting for follow-up time

Students should have some familiarity with the different ways of expressing exposure. Exposure can be described in terms of person-time of exposure or as counts of people exposed versus non-exposed. Each method is based on different assumptions. It is helpful for the students to grasp these differences by giving examples. In order that use of person counts be a valid approach, there must be comparable follow-up for all study subjects. This would apply in the study of an acute effect of an antibiotic, for example, where each person is followed for only forty-five days. However, in a study of a chronically used drug, use will stop, start, change over time, and vary from one person to another. In this situation, the contribution of each individual can be accumulated using person-time.

Considerations for case-control studies

Selection of cases and controls

In any drug safety study it is important to select cases with a newly diagnosed illness that has no other apparent non-drug cause. One is not interested in studying prevalent disease. Said another way, if a disease is present prior to receipt of the drug, then there is no possibility that the drug caused the disease and therefore that person should not be included as a case. Similarly, if a subject has another proximate cause present for the illness of interest that is the most likely cause of the outcome, then that person should not be included as a case (H. Jick and Vessey 1978). For example, in the case of a person who has just had surgery and develops a venous thromboembolism, the illness is most likely a result of the surgery and the drug should not be considered as a possible cause. Here, again, it is helpful to demonstrate to the students the importance of this lesson by providing examples and showing how the results differ when all cases are included, instead of only idiopathic cases (see Table 10.1). Case selection should always be accomplished without knowledge of the subject's exposure status in order to avoid biased case selection.

When selecting controls, any exclusion criteria applied to the cases must also be applied to the controls. In other words, the base population from which the controls are drawn should be properly selected so that the cases and controls are equally likely to be exposed to the study drug *in the absence of a drug effect*. For example, if studying oestrogens and endometrial cancer, all women in the study population from whence cases and controls are drawn should have an intact

uterus. A woman without a uterus could not become a case, nor does she have the same likelihood of being exposed to oestrogens.

It is often helpful to match controls to cases on age, sex, and location to facilitate the control of these potential confounding factors. In addition, matching the index date of the case to a set of matched controls achieves control of calendar time in an efficient manner. It should be noted, however, that once matched on a factor, the independent effects of that factor cannot be evaluated in the analysis, and the analysis must take the matching into account.

Selection of controls in a drug safety study should be discussed in depth with the students. It is helpful to go through several examples to illustrate the principles involved here. If one is selecting controls from a hospital population, it is important that the controls not be admitted to the hospital for a reason related to the outcome or the exposure under study. For example, in a study of oral contraceptives and breast cancer, it is important that controls not be selected from a cardiovascular ward, as these women are *less* likely to be taking oral contraceptives since women with cardiovascular problems are selectively not prescribed the drug. Selecting controls from this population would bias the likelihood of exposure in a way that could create a spurious positive association between oral contraceptives and breast cancer. Similarly, it is important that the controls not be *more* likely to be exposed to the drug of interest. For example, if one were studying the association between antidepressants and convulsions, one would not select controls from the psychiatric ward, as they would be more likely to be taking antidepressant medications. It would be appropriate to select controls from among those who were admitted for a condition that was unrelated to any psychiatric or neurological problem (see H. Jick and Vessey 1978 for a more thorough discussion of this topic).

Expressing exposure

In a case-control study, exposure is assessed for each case and each control for the period of time prior to the index date. Subjects are categorized into a predefined set of exposure categories such as current user, past user, non-user, etc. Exposure can then be further divided into categories of duration of use or dose. It is useful to discuss how to define the different exposure, dose, and duration categories, and to illustrate how the definitions would change for different drug-disease relations. For example, in studies of antibiotics, most use is very short duration; however, it would be important to consider different daily doses, since some adverse events are strongly dose related. In a study of an NSAID in relation to cancer, one should look at the duration of use, the number of prescriptions received (since one does not necessarily use NSAIDs continuously and this would capture how much the NSAID was used during the total

duration of use), and daily dose of the NSAID. Total cumulative dose could also be assessed, as well as time from first and last NSAID to the index which may also be important in this type of study. A study by Khurana et al. (2007) of statin use in relation to the risk of lung cancer provides an instructive example of the importance of controlling for calendar time when assessing exposure by, for example, assigning the index date of each case to the matched controls. In this study each case's exposure was determined based on exposure prior to the index date (date of diagnosis), while, for each control, all exposure up to the end of the patient record was used to determine exposure status. As a result, controls had longer exposure opportunity than cases and therefore more exposure, an observation that explains the resulting protective effect of statins on the risk for lung cancer. This is another example that illustrates the importance of careful exposure definition and ascertainment.

Defining exposure windows

Exposure windows in case-control studies have the same considerations as those in cohort studies. Depending on the outcome under study, one may be interested only in exposures that occurred in the forty-five or sixty days prior to the index date, or in the time years prior to the index date; alternatively, one may be interested in knowing about all study drug exposure prior to the index date.

Drugs in relation to congenital anomalies

The study of congenital anomalies in relation to drug exposures has special considerations. A congenital anomaly, strictly speaking, is a prevalent condition. However, it remains true that one is interested in studying exposure to drugs that occurs prior to the development of the anomaly in utero. If a woman takes a drug in the third trimester of pregnancy, after the foetus has developed, that drug could not be responsible for causing spina bifida, for example, as the neural tube had already been formed at the time of exposure. In general, one would like to know of all exposures that occurred just prior to conception and during the first trimester of pregnancy, when the development of the body organs and limbs occurs. The exposure that occurs at the time the anomaly becomes manifest (at birth) is not relevant to the etiology of the outcome. This can pose a problem for ascertaining accurate exposure information, since the exposure occurred long before the outcome was known. The issues here are not unlike the issues present in studies of cancer and other chronic diseases where there is a delay between exposure and diagnosis of disease. A discussion of which data resources would best be suited for these studies is helpful in thinking through these issues.

Data resources

I like to describe and discuss with my students some of the resources that are available for conducting drug safety research. Today, there are many resources available but below are some that have been used repeatedly and have been shown to be of high quality for this kind of research.

Ad hoc studies

Some studies are designed to address safety issues for a specific drug or set of drugs. The Oxford Family Planning Association Contraceptive Study is one such example. In this study, there was interest in evaluating the long-term effects of oral contraceptive pills. A group of women who took oral contraceptives was followed over a long period of time and compared to a group of women who used other contraceptive methods. The study required identifying and interviewing many women and then following them and repeatedly interviewing them over time to maintain accurate records of exposure and outcome information (Vessey 1998). Other similar studies exist but there are few that were designed specifically to study drug effects.

Electronic and administrative data resources

In the past three decades the use of automated or administrative data for conducting drug safety studies has increased greatly as more electronic data resources have become available (H. Jick 1985). In the late 1970s data from Group Health Cooperative of Puget Sound in Seattle, Washington, was first used to study drug/disease associations. Other automated databases, including the Saskatchewan Health database, other health maintenance organization data resources such as the Kaiser Permanente database and the Health Maintenance Organization Research Network database, the Clinical Practice Research Datalink (formerly known as the General Practice Research Database), and various Medicaid/Medicare databases, among others, have since been used regularly for the study of drug effects (H. Jick et al. 1984; Strand 1985; Ray and Griffin 1989; H. Jick et al. 1991; Malcolm et al. 1993; Lanza et al. 1995; also see <http://www.cprd.com>). US based insurance claims databases including PharMetrics/IMS and MarketScan have also seen increasing use in the area of drug safety research (S. Jick et al. 2006). In addition, government data resources from Europe have been used more frequently; some examples of these include the Dutch Integrated Primary Care Information Database (ICPI) and the Danish registries (Sorensen and Larsen 1994; Vlug et al. 1999; Ehrenstein et al. 2010). The advantage of these databases is that they contain information prospectively recorded on virtually all prescriptions used by individuals covered in the

database. As a consequence, one is not dependent on the expensive and time-consuming job of recording by hand all drugs used by an individual, as in a prospective paper-based study, nor is one relying on subject recall of drug use (as in a retrospective study), which may be of questionable validity (Klemetti and Saxen 1967). These databases also have complete information on outcomes. In some databases there are only hospitalized outcomes, while others have both inpatient and outpatient diagnoses. Finally, these databases cover many patients, and previous limitations on sample size are now less of a concern.

Some data resources include long-term follow-up of patients and are thus appropriate for use in the study of delayed outcomes, while other resources have short average follow-up and thus their utility is limited to studies of acute outcomes. There are other important differences between the databases that must be considered when choosing the 'best' data resource for a particular question. For example, access to the original clinical records for purposes of data validation is possible in some data resources but not others. Another important difference to consider is the motivation for data recording. In the Clinical Practice Research Datalink (CPRD) database, the data comprises the physicians' medical record for each patient; thus physicians are motivated to keep complete and accurate records. In contrast, in claims databases information is entered based on reimbursement considerations, not for the sake of keeping a complete and accurate medical record. These factors can have an effect on the accuracy and completeness of the information present in the database; that is, accuracy of outcome recording differs between databases. Other considerations include the availability of the drug of interest in each database, and age restrictions that may have implications for the study in question. Some databases are restricted by a formulary in the insurance plan or by the country of origin. The size of the database, the frequency of use of the drug of interest, as well as the frequency of the outcome are all additional considerations in database selection. While not all databases are appropriate for all studies, the availability of these data resources has revolutionized research in this area, making studies less expensive and faster to complete, and allows the conduct of studies that were not previously feasible.

Analysis and interpretation of results

If a study has been designed properly, the analysis is often straightforward. One must always control for confounding, as discussed in 'Control of confounding' and 'Confounding by indication', even if it has been controlled in the study design by matching. It is often important to consider dose and duration of use in the analysis of a study. If there is only a slight increase in the overall risk of an

outcome relative to a drug, it is possible that the risk is limited to those in a subgroup of the exposed, such as those with long-term use, high-dose exposure, or some other factor. Other subgroups should also be evaluated when relevant, such as older subjects (or younger), etc.

If a positive association is found between a drug and an exposure, one has to ask if the association is causal or if it is due to some other factor(s) such as confounding, bias including confounding by indication, poor study design, or improper analysis. Students should understand the particular difficulty of assessing causality in conducting drug safety studies, where there is often a question as to the role of the underlying disease being treated. Also of importance are questions of comparison group selection. Was the comparison group chosen properly to avoid selection bias? Was exposure assessed in the same way for the cases and controls? Were the exposed and non-exposed defined properly? Was the analysis properly conducted? Is there another explanation for the association that is not causal? For example, in a study of inhaled corticosteroids and the development of cataracts, could an association with heavier use and cataract be due to the drug or could it be that more severe asthma is itself associated with increased risk for cataract? Students should think through these possible explanations carefully.

As with all epidemiologic studies, a null result must also be interpreted with the appropriate considerations, including misclassification of exposure, negative confounding, or as mentioned above, misclassification of the outcome. In drug studies in particular, the source of information on exposure can be more or less subject to misclassification. It is helpful to have the students think through the effects of biased and random misclassification of exposure. For example, if you have incomplete exposure information because you only have information obtained at one interview five years prior to the outcome under study, it is likely that many people who were exposed at that time have stopped using the drug under study and that others who were unexposed at interview have since taken the drug. This would result in some people being misclassified as unexposed and some people who may have been exposed only long ago being incorrectly considered exposed at the index date.

Assessing students' achievements

Below are some exercises, the product of which provides the teacher with good information on a student's grasp of the material covered in the course. One such exercise is to provide the students with several published studies on the same topic for which the result is not consistent across the studies, such as, for example, the risk of venous thromboembolism in users of oral contraceptives

(e.g. Dinger et al. 2007; Seeger et al. 2007; Lidegaard et al. 2009; van Hylckama et al. 2009; S. Jick and Hernandez 2011) or the use of COX-2 inhibitors and the risk of myocardial infarction (e.g. Bresalier et al. 2005; Graham et al. 2005; Hippisley-Cox and Coupland 2005; Andersohn et al. 2006; H. Jick et al. 2006). Have the students review each article and compare the critical methodologic components across the studies: study design, study population, exclusion criteria, exposure and outcome definitions, selection of cases/controls or exposed/non-exposed, data resource, control of confounding, analysis, results, and interpretation. The students who more thoroughly grasp the material will uncover more subtle and critical differences in the studies, while others will demonstrate a more superficial understanding of the findings.

Some of the key differences in the venous thromboembolism studies relate to the selection, validation, and definition of cases, particularly the inclusion of non-idiopathic cases in some studies. See the paper by Stolley et al. 1975 for an excellent demonstration of the importance of excluding non-idiopathic cases. In the COX-2 studies, differences are present in both the definition of the outcome (any myocardial infarction, vs any incident myocardial infarction, vs idiopathic myocardial infarction, vs non-idiopathic myocardial infarction) and in the definition of exposure. The exposure definitions vary greatly in relation to the relevant exposure windows: current, recent, and ever-use, definition of the reference group, and the evaluation of duration and dose. The implications of these study design differences on the different results should be explored and discussed.

Another exercise is to provide a study topic and ask students to select the most appropriate database (and the reasons why) and to design a study to best answer the question at hand. Students should indicate why they chose their particular study design, including case, exposure and referent definitions, and data resource. Possible topics could include studies of NSAIDs and renal stones, cholesterol-lowering agents and cancer, oral hypoglycaemic drugs and liver disease, or COX-2 inhibitors and gastrointestinal bleed. Each study has its own methodologic complexities that will challenge the students to employ the techniques taught in class. Selection of the appropriate comparison groups is always critical in these studies to control for confounding by indication.

References

Andersohn, F., Suissa, S., and Garbe, E. (2006) Use of first- and second-generation cyclooxygenase-2-selective nonsteroidal antiinflammatory drugs and risk of acute myocardial infarction. *Circulation*, **113**: 1950–7.

Avorn, J. (2007) Keeping science on top in drug evaluation. *New England Journal of Medicine*, **357**: 633–5.

Boston Collaborative Drug Surveillance Program. (1974) Regular aspirin intake and acute myocardial infarction. *British Medical Journal*, **1**: 440–3.

Bresalier, R. S. et al. (2005) Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *New England Journal of Medicine*, **352**: 1092–102.

Curfman, G. D., Morrissey, S., and Drazen, J. M. (2007) Safer drugs for the American people. *New England Journal of Medicine*, **357**: 602–3.

Derby, L., Jick, H., Henry, D. A., and Dean, A. D. (1993) Cholestatic hepatitis associated with flucloxacillin. *Medical Journal of Australia*, **158**: 600–2.

Dinger, J. C., Heinemann, L. A., and Kühl-Habich, D. (2007) The safety of a drospirenone-containing oral contraceptive: final results from the European Active Surveillance Study on oral contraceptives based on 142,475 women-years of observation. *Contraception*, **75**: 344–54.

Ehrenstein, V., Antonsen, S., and Pedersen, L. (2010) Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clinical Epidemiology*, **2**: 273–9.

García Rodríguez, L. A. and Jick, H. (1994) Comparison of the risk of gynaecomastia associated with cimetidine, omeprazole and other antiulcer medications. *British Medical Journal*, **308**: 503–6.

Graham, D. J., Campen, D., Hui, R., Spence, M., Cheetham, C., Levy, G., Shoor, S., and Ray, W. A. (2005) Risk of myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. *Lancet*, **365**: 475–81.

Gray, S. L., LaCroix, A. Z., Larson, J., Robbins, J., Cauley, J. A., Manson, J. E., Chen, Z. (2010) Proton pump inhibitor use, hip fracture, and change in bone mineral density in postmenopausal women; Results from the Women's Health Initiative. *Archives of Internal Medicine*, **170**: 765–71.

Hampton, T. (2007) Postmarket 'Pharmacovigilance' program on alert for adverse events from drugs. *Journal of the American Medical Association*, **298**: 851–2.

Hippisley-Cox, J. and Coupland, C. (2005) Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. *British Medical Journal*, **330**: 1366–72.

Jick, H. (1977) The discovery of drug-induced illness. *New England Journal of Medicine*, **296**: 481–5.

Jick, H. (1985) Use of automated data bases to study drug effects after marketing. *Pharmacotherapy*, **5**: 278–9.

Jick, H., García Rodríguez, L. A., and Pérez Gutthann, S. (1998) Principles of epidemiological research on adverse and beneficial drug effects. *Lancet*, **352**: 1767–70.

Jick, H., Jick, S. S., and Derby, L. E. (1991) Validation of a large general practice based database in the UK. *British Medical Journal*, **302**: 766–8.

Jick, H., Kaye, J. A., Russmann, S., and Jick, S. S. (2006) Nonsteroidal antiinflammatory drugs and acute myocardial infarction in patients with no major risk factors. *Pharmacotherapy*, **26**: 1379–87.

Jick, H., Madsen, S., and Nudehnan, P. M. (1984) Postmarketing follow-up at Group Health Cooperative of Puget Sound. *Pharmacotherapy*, **4**: 99–100.

Jick, H., Miettinen, O. S., Shapiro, S., Lewis, G. P., Siskind, V., and Slone, D. (1970) Comprehensive drug surveillance. *Journal of the American Medical Association*, **213**: 1455–60.

Jick, H. and Vessey, M. P. (1978) Case-control studies in the evaluation of drug induced illness. *American Journal of Epidemiology*, **107**: 1–7.

Jick, H., Watkins, R. N., Hunter, J. R., Dinan, B. J., Madsen, S., Rothman, K. J., and Walker, A. M. (1979) Replacement estrogens and endometrial cancer. *New England Journal of Medicine*, **300**: 218–22.

Jick, S. S., Hagberg, K. W., Kaye, J. A., and Jick H. (2009) Postmenopausal estrogen-containing hormone therapy and the risk of breast cancer. *Obstetrics and Gynecology*, **113**: 74–90.

Jick, S. S. and Hernandez, R. K. (2011) Risk of non-fatal venous thromboembolism in women using oral contraceptives containing drospirenone compared with women using oral contraceptives containing levonorgestrel: case-control study using United States claims data. *British Medical Journal*, **340**: d2151.

Jick, S. S., Kaye, J. A., Russmann S., and Jick H. (2006) Risk of non-fatal venous thromboembolism in women using a contraceptive transdermal patch and oral contraceptives containing norgestimate and 35 ug of ethinyl estradiol. *Contraception*, **73**: 223–8.

Jick, S. S.,Vasilakis, C., Martinez, C., and Jick H. (1998) A study of the relation of exposure to quinolones and suicidal behaviour. *British Journal of Clinical Pharmacology*, **45**: 77–81.

Jick, S. S., Vasilakis-Scaramozza, C., and Maier, W. C. (2001) The risk of cataract among users of inhaled steroids. *Epidemiology*, **12**: 229–34.

Kaye, J. A. and Jick, H. (2008) Proton pump inhibitors and hip fractures in patients without major risk factors. *Pharmacotherapy*, **28**: 951–9.

Khurana, V., Bejjanki, H. R., Caldito, G., and Owens, M. W. (2007) Statins reduce the risk of lung cancer in humans: a large case-control study of US veterans. Chest, **131**: 1282–8.

Klemetti, A. and Saxen, L. (1967) Prospective versus retrospective approach in the search for environmental causes and malformations. *American Journal of Public Health*, **57**: 2071–5.

Lanza, L. L., Walker, A. M., Bortnichak, E. A., and Dreyer, N. A. (1995) Peptic ulcer and gastrointestinal hemorrhage associated with nonsteroidal anti-inflammatory drug use in patients younger than 65 years: a large health maintenance organization cohort study. *Archives of Internal Medicine*, **155**: 1371–7.

Lidegaard, Ø., Løkkegaard, E., Svendsen, A. L., and Agger, C. (2009) Hormonal contraception and risk of venous thromboembolism: national follow-up study. *British Medical Journal*, **339**: b2890.

Malcolm, E., Downey, W., Strand, L. M., McNutt, M., and West, R. (1993) Saskatchewan Health's linkable databases and pharmacoepidemiology. *Postmarketing Surveillance*, **6**: 175–264.

Ray, W. A. (1988) Pharmacoepidemiology: is ignorance bliss? *Journal of Clinical Research and Drug Development*, **2**: 67–74.

Ray, W. A. and Griffin, M. R. (1989) Use of Medicaid data for pharmacoepidemiology. *American Journal of Epidemiology*, **129**: 837–49.

Rosen, C. J. (2007) The rosiglitazone story—lessons from an FDA advisory committee meeting. *New England Journal of Medicine*, **357**: 844–6.

Schumacher, M., Jick, S. S., Jick, H., and Feld, A. D. (1990) Cimetidine use and gastric cancer. *Epidemiology*, **1**: 251–4.

Seeger, J. D., Loughlin, J., Eng, P. M., Clifford, C. R., Cutone, J., and Walker, A. M. (2007) Risk of thromboembolism in women taking ethinylestradiol/drospirenone and other oral contraceptives. *Obstetrics and Gynecology*, **110**: 587–93.

Sorensen, H. T. and Larsen, B. O. (1994) A population-based Danish data resource with possible high validity in pharmacoepidemiological research. *Journal of Medical Systems*, **18**: 33–8.

Stolley, P. D., Tonascia, J. A., Tockman, M. S., Sartwell, P. E., Rutledge, A. H., and Jacobs, M. P. (1975) Thrombosis with low-estrogen oral contraceptives. *American Journal of Epidemiology*, **102**: 197–208.

Strand, L. M. (1985) Drug epidemiology resources and studies: the Saskatchewan database. *Drug Information Journal*, **19**: 253–6.

Tamimi, R. M., Hankinson, S. E., Chen, W. Y., Rosner, B., and Colditz, G. A. (2006) Combined estrogen and testosterone use and risk of breast cancer in postmenopausal women. *Archives of Internal Medicine*, **166**: 1483–9.

van Hylckama Vlieg, A., Helmerhorst, F. M., Vandenbroucke, J. P., Doggen, C. J. M., and Rosendaal, F. R. (2009) The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *British Medical Journal*, **339**: b2921.

Vessey, M. (1998) 30th Anniversary of the Oxford-FPA Contraceptive Study. *Trends in Urology, Gynaecology and Sexual Health*, **3**: 26–33.

Vlug, A. E., van der Lei, J., Mosseveld, B. M., van Wijk, M. A., van der Linden, P. D., Sturkenboom, M. C., and van Bemmel, J. H. (1999) Postmarketing surveillance based on electronic patient records: the IPCI project. *Methods of Information in Medicine*, **38**: 339–44.

Walker, A. M. (1996) Confounding by indication. *Epidemiology*, **7**: 335–6.

Yang, Y. X., Lewis, J. D., Epstein, S., and Metz, D. C. (2006) Long-term proton pump inhibitor therapy and the risk of hip fracture. *Journal of the American Medical Society*, **296**: 2947–53.

Chapter 11

Nutritional epidemiology

Walter C. Willett

Introduction to nutritional epidemiology

Diet has long been thought to have important impacts on human health. Roughly one-third of cancers and an even higher percentage of cardiovascular diseases are thought to be diet related but the specific aspects of diet responsible have been less clear. Many other conditions previously not thought to be diet related, including birth defects, cataracts, renal stones, and other degenerative conditions, have also been found to have dietary etiologies. Nutritional epidemiology is a relatively new branch of epidemiology and focuses on understanding the relation between diet and long-term health and disease. Our understanding of biologic mechanisms remains far too incomplete to predict confidently the ultimate consequences of eating a particular food or nutrient. Therefore, epidemiologic studies directly relating intake of various dietary components to risk of disease among humans, complemented by laboratory investigations and mechanistic studies, will play a critical role in guiding individual food choices and public policy.

In the 1980s a major expansion of the literature in nutritional epidemiology occurred and a firmer quantitative basis for this field developed. In particular, the substantial variation in diet among individuals was quantified in many populations, standardized dietary questionnaires were developed for large epidemiologic studies, and the ability of these questionnaires to measure diet was documented. Many of the major questions about diet and disease remain unresolved but the foundations for obtaining such information became relatively well established. Thus, a strong basis was created for formal courses on the topic of nutritional epidemiology.

Any teaching activity should take into account the background and interests of the students. The basic nutritional epidemiology course that we teach assumes students have already taken introductory epidemiology and biostatistics courses or have had equivalent experience. The depth is greater than would

usually be appropriate for medical students or students pursuing a master's degree in public health, unless they have a special interest in this topic.

Among the various audiences are researchers actively engaged in studies of diet and disease, and other persons who are attempting to read and interpret published epidemiologic reports related to nutrition. Many epidemiology students whose primary interest is not nutritional epidemiology often take our course because diet needs to be considered in almost any field, if nothing more than as a potential confounding variable or effect modifier. Potential students may include those who are simply interested in the most up-to-date knowledge about diet and health. In general, I discourage such students from taking a methodologic course in nutritional epidemiology, as the substantive knowledge is changing rapidly and an emphasis is therefore appropriately given to fundamental methodologic issues. The course I have taught and will describe here specifically does not address problems of undernutrition in developing countries. That topic has a long tradition, mainly based on anthropometric measurements, and is covered elsewhere in our School of Public Health. Nevertheless, students interested in such problems can gain much by a general course in nutritional epidemiology because the current methods can be applied to problems of undernutrition in children.

In the early 1980s a single course could almost exhaustively cover the area of nutritional epidemiology because the literature was limited. However, as in any field, further development moves the methodologic frontier more distant from the level of knowledge needed by someone entering the field. Thus, teaching a course that serves both as an introduction and as an up-to-date account of new developments becomes increasingly challenging. For this reason, we are now teaching both an introductory basic course for those entering the field as well as an advanced level course that covers emerging methodologic issues in more detail.

Many epidemiology students find nutritional epidemiology to be useful in a wider epidemiologic context because many of the issues are not traditionally covered well in basic epidemiology curricula. For example, most exposure measurements in epidemiology courses are considered as categorical variables, whereas in nutrition most exposure variables, as well as many outcome variables, are continuous. As another example, measurement error and corrections for measurement error are generally not included in our primary epidemiology curriculum. As much of the work in this area has emanated from nutritional epidemiologic problems, this topic is covered in our course but the principles have wide application throughout epidemiology. Thus, one of the course objectives is to reinforce epidemiologic principles that apply to other areas of our broader field.

Teaching objectives

Students who complete our introductory course in nutritional epidemiology should

1. be familiar with the basic concepts in this field;
2. be able to read and interpret critically published articles relating diet to disease; and
3. be able to conduct basic analyses using dietary data, including adjustment for total energy intake and de-attenuation of correlation coefficients, and interpret findings from both a statistical and biological perspective.

Format

Our course in nutritional epidemiology consists of four major components: lectures, readings, data collection experiences, and computer problem sets. The mix of these would appropriately depend on the time allotment and student background. Ordinarily, the introductory course is given as sixteen two-hour periods, plus individual work on data collection and computer analysis. About half of the lecture sessions are devoted to methodologic issues in the measurement and analysis of dietary data and are illustrated with relevant examples. The second half of the course uses substantive topics in nutritional epidemiology to reinforce the principles.

As usual, a lecture format is optimal when the class size is sufficiently small to allow interaction when students have done preparatory readings. For our teaching, we have used my textbook, *Nutritional Epidemiology* (Willett 2013), as the primary material, which is supplemented by readings of original reports. For the substantive chapters, which rapidly become out of date, the book is used as a summary of thinking on a topic as of 2013, and more recent papers are used to illustrate the evolution of insights. The other major textbook in this field is by Margetts and Nelson (1997).

Data collection by students has involved the keeping of multiple 24-hour recalls, diet records, and the completion of a food frequency questionnaire. Students enter and analyse their own diet record and 24-hour recall data using nutrient analysis software (currently we use a free online programme (<http://www.nutritiondata.com/>)). The food frequency questionnaire is scanned and analysed by our standard procedures. Individual and summary reports for the class are returned to the students and serve as the basis for a discussion about sources of error.

The computer problem sets provide students with actual experience in manipulating nutritional data; the topics relate to specific issues that are discussed in the

lectures and readings. Typically, students have found this to be one of the most valuable aspects of the course, as this requires an understanding of the methodological issues and provides students confidence that they are able to analyse information appropriately. The major problems deal with the analysis of within- and between-person components of variation in diet, validation of a food frequency questionnaire, examination of the relation between dietary intake and blood levels of nutrients, and correction of correlation coefficients for measurement error. These will be discussed further in the relevant topics below. Students are usually required to turn in the primary computer output and a discussion of the results incorporating both the biological and statistical considerations.

Teaching content

In the following sections, the major lecture topics will be discussed and the most important points noted. The abridged syllabus is shown in Table 11.1.

Table 11.1 Syllabus for introductory nutritional epidemiology

Session 1	Overview of epidemiology in nutrition studies Sources of variation in the diet (chapters 1 and 3) Introduction to computer assignments (Hand out homework 1) Readings: Beaton et al. (1979)
Session 2	Dietary methodology: food frequency (chapters 2 and 5) Dietary methodology: validation (chapter 6) Dietary assessment assignment Readings: Willett et al. (1985)
Session 3	Body composition/anthropometry (chapter 9) Obesity and mortality Readings: Willett et al. (1999); Flegal et al. (2005)
Session 4	Dietary methodology: 24-hour recalls and diet records (chapter 4) (Homework 1 due; hand out homework 2)
Session 5	Meaning and analysis of total energy intake (chapter 11) Readings: Willett et al. (1997)
Session 6	Dietary methodology: nutrient database (Homework 2 due; hand out homework 3)
Session 7	Introduction to regression analysis of nutrient Review of homeworks 1 and 2
Session 8	Correction for measurement error (chapter 12)
Session 9	Biochemical assessment of nutritional status (chapter 8)
Session 10	Data analysis and presentation in nutritional epidemiology (chapter 13)
Session 11	Folic acid and neural tube defects (chapter 20) (Homework 3 due; hand out homework 4)
Session 12	Diet and heart disease (chapter 19) Readings: Shekelle et al. (1981); Hu et al. (1997)

Table 11.1 (continued) Syllabus for introductory nutritional epidemiology

Session 13	Diet and breast cancer (chapter 18) (Diet assessment due) Readings: Kim et al. (2006)
Session 14	Vitamin A and cancer (chapter 17) (Homework 4 due) Readings: Hennekens et al. (1996); ATBC (1994); Mannisto et al. (2004)
Session 15	Review homeworks 3 and 4 Review of course
Session 16	Examination

Overview

Nutritional epidemiology is introduced as essentially a subdivision of epidemiology, with diet as the exposure and disease as the outcome. However, this field is unusually complex because of the many dimensions of diet representing many correlated variables and likely interactions with other nutrients, genetic factors, obesity, and physical activity.

The first session is devoted to an overview of nutritional epidemiology. A brief historical background is provided. The historic role of international ecological studies in nutritional epidemiology is emphasized, as well as the potentially large and often intractable problems of confounding in such studies. The further development of nutritional epidemiology was hindered by the conventional wisdom that there was no heterogeneity in diet within populations and that individuals could not remember what they had previously eaten. A large amount of this negativity arose because of the inability to document correlations between dietary intake of cholesterol and serum cholesterol. There are several reasons for this minimal correlation, the most important being that serum cholesterol is heavily controlled by homeostatic factors, and a substantial part of the between-person variation appears to be under genetic control. Thus, at most, very weak associations between diet and serum cholesterol should be expected. Furthermore, most studies of diet had used only a single 24-hour recall, which necessarily provides a poor assessment of usual intake because of large day-to-day variation. Thus, the focus on serum cholesterol as the criteria for validity of dietary assessment methods was highly unfortunate.

In addition to ecological studies, the importance of migrant studies and secular trends in the development of nutritional epidemiology is emphasized. Importantly, such studies indicate that the primary determinants of disease rates of cancer and cardiovascular disease are primarily due to non-genetic factors. The role of case-control and cohort studies in nutritional epidemiology is also discussed. The problem of bias in case-control studies of diet is particularly

substantial because of the modest relative risks that are to be expected and the sensitivity of relative risks to even small degrees of bias due to differences in recall between cases and controls or selection of unrepresentative controls. Cohort studies are primarily limited by feasibility because of the large sizes and the long duration of follow-up that are typically needed.

Uncertainty about temporal relationships between exposures and occurrence of disease poses great challenges to epidemiology, including nutritional epidemiology. For example, evidence that birth weight (Michels and Xue 2006), adiposity at ages five and ten (Baer 2005), and rate of height gain during adolescence are associated with breast cancer incidence suggests effects of nutrition before birth and during childhood; however, associations with weight gain and loss after menopause (Eliassen 2006) also indicate that nutritional factors can act late in life. Rarely can a single study address the potential effects of diet throughout the life cycle. Rather, evidence will usually need to be pieced together from studies beginning before birth and during childhood in relation to intermediate risk factors, studies of adults using data on diet during childhood assessed retrospectively but with disease incidence evaluated prospectively (Linos et al. 2008), and prospective studies of adults enrolled at various ages (Kim et al. 2006). Analytic approaches to investigate different temporal relationships are discussed in session 10 of the lectures.

The role of controlled trials in nutritional epidemiology is also discussed. Again, feasibility is a major issue, in part because of the problem of maintaining high levels of compliance over long periods of time, uncertainty about the relevant follow-up period, and the large sample sizes required to test most hypotheses. Because of the impossibility of testing most diet and disease hypotheses in randomized trials, the importance of sound observational study design is obvious.

Food and nutrients

In this section, the complexity of human diets is examined. The foods we consume each day contain thousands of specific chemicals, some of which are known and well quantified, and others that are poorly described or not even measured. These dimensions of diets include the essential nutrients (including minerals, vitamins, lipids, and amino acids), major energy sources (proteins, carbohydrates, fats, and alcohol), additives (nitrates, salts, colouring agents, etc.), agricultural contaminants (pesticides, herbicides, fungicides, and growth hormones), microbial toxin contaminants (aflatoxin being a classic example), inorganic contaminants (including lead, cadmium, and PCBs), chemicals formed in the cooking or processing of food (various mutagens, *trans* fatty

acids for example), natural toxins (including a plethora of phytochemicals that plants have developed through evolution to deter insects and diseases), and other natural compounds (for example, the constituents of normal cell structures). Available databases allow intakes to be estimated for only a small fraction of these various constituents of foods, mainly the essential nutrients and energy sources.

The value of examining both foods and nutrients in relation to disease risks is addressed. If a specific nutrient is a cause or preventive factor for disease, the most powerful relationship will be achieved by examining this nutrient. Nevertheless, confidence in causality can be increased if major food sources of that nutrient are also similarly related to risk of disease. On the other hand, an examination of foods may be revealing because the important dietary constituent may not be represented by any calculable variable. Moreover, foods are always complex mixtures and their relationship to disease risk cannot necessarily be predicted by examination of a single constituent. Because dietary advice will often be made on the basis of foods, it will be important to examine intake of specific foods directly in relation to disease risk. Thus, there is great value in examining foods and nutrients simultaneously in relation to disease outcomes.

Food composition data sources and computation systems

The calculation of nutrient intakes from data on food consumption requires a food composition database. The underlying assumption in this calculation is that the nutrient content of a specific food does not vary from one sample to another. Naturally, this is never completely true because the composition will depend on many factors, including the growing and harvesting conditions, the variety, the degree of maturity, and factors in the processing, storage, and cooking. In most cases this is not a serious problem; however, in some instances, the variability may be so great that calculations become useless. Selenium is a classic example, as the content in food can vary over a hundred-fold, depending on the soil in which the food was produced.

Considerations in selection of databases are reviewed in this session of our course. Key points are that the food composition be as accurate and up to date as possible, that uniformity in the determination in nutrient composition across foods is desirable for each specific nutrient, that the database is comprehensive in the scope of foods because all foods reported must be assigned nutrient values, that the specificity is adequate for nutrients that will be evaluated (for example, the composition of margarines can vary dramatically, depending on

the type of oil used and the processing), and that the range of nutrients included should be as comprehensive as possible. In the US, the Department of Agriculture maintains the most comprehensive system overall but, for many dietary variables, this will need to be supplemented with further information. For other countries, investigators will need to determine the most appropriate food composition sources.

In addition to a food composition database, computer software will be needed for the computation of nutrient intakes from dietary data. For food frequency questionnaires, investigators will generally need to assemble their own corresponding database to match the foods that are included on their questionnaire. For more open-ended methods such as 24-hour recalls or dietary records (see 'Short-term dietary assessment methods: 24-hour recall and food records'), more extensive database systems are required. Fortunately, a wide variety of options are now available that can aid in the efficient analysis of dietary information.

Nature of variation in diet

An understanding of the sources of variation in dietary intake is essential for nutritional epidemiologists. In addition to the relevant chapter in *Nutritional Epidemiology*, students are asked to read the classic paper by Beaton et al. (1979) on this topic.

For most individuals, nutrient intake varies tremendously from day to day. In some circumstances, particularly in developing countries where food availability changes dramatically by season, this can also be a major source of variation. The most important implication is that a single 24-hour recording of food intake, no matter how accurate, will generally be a poor representation of a person's average long-term intake. Moreover, the degree of variability differs dramatically from one nutrient to another. Only for total energy intake is there potent physiologic regulation to dampen day-to-day variation. Thus, variability is lowest for total energy intake and next lowest for the major contributors to energy intake, specifically, total fat, carbohydrate, and protein. Minor constituents of the diet are much less constrained; for example, the day-to-day variability in vitamin A or cholesterol can be extremely large. Because the absolute intake of nutrients is in part determined by total energy intake, the composition of diets (intakes adjusted for energy intake, discussed further in 'Implications of total energy intake for epidemiologic analyses') can be considerably more variable than for absolute intake.

The implications of large day-to-day variability for epidemiologic studies are also reviewed. These include that observed distributions based on single 24-hour

recalls are much broader than the true distributions of long-term individual intakes. Also, measures of association such as relative risks, and correlation and regression coefficients are biased toward the null. As part of this topic, students receive a problem set with multiple days of 24-hour intake and conduct an analysis of variance to partition the within-person and between-person components of variance for several nutrients.

Short-term dietary assessment methods: 24-hour recall and food records

In this section the process of collecting 24-hour recall and diet record data is described by a dietician experienced in these methodologies. The advantages and disadvantages of 24-hour recall and diet record methods are reviewed, a particular advantage being that there is no constraint placed on the food data and no assumptions made about ways of consuming foods or portion sizes. A strength of the 24-hour recall method is that it does not require literacy or any substantial effort on the part of participants. On the other hand, an advantage of food records is that this method does not depend on memory, and quantities can actually be measured and weighed directly at the time of recording. However, diet record collection requires a highly motivated and literate participant.

For both of these methods, the effort in collection and processing of dietary data is large, and this has generally precluded these methods in large prospective studies. Recently, web-based 24-hour recalls have become available, which greatly reduce the cost and may make this methodology useful in a broader range of applications. These short-term methods only provide information about current diet and thus may not be relevant in case-control studies, where past diet is usually of interest. A major role of 24-hour recalls and diet records in nutritional epidemiology is for the validation of food frequency questionnaires. Also, for description of group means or for cross-cultural comparisons, 24-hour recall information may be optimal.

Food frequency methods

Because of the practical and conceptual limitations of 24-hour recalls and food records for assessing average long-term intake, most investigators have converged upon the use of food frequency questionnaires in epidemiologic studies. The underlying principles of this method were described by the British statistician Heady, who noted that differences in nutrient intakes were primarily determined by the frequency with which foods were consumed rather than differences in serving sizes (Heady 1961). In addition, cognitive research has

documented that reporting of usual food consumption is generally easier than describing what was eaten at a specific meal.

The design of food frequency questionnaires is considered. Decisions in creating a food list are reviewed, with the most important criteria being that foods most accounting for between-person variation in nutrients of interest should be given priority. A variety of ways to identify foods for inclusion on a food frequency questionnaire are described. Decisions about the collection of frequency information are also considered, including whether open-ended responses or a multiple-choice format is preferable. The multiple-choice format has major advantages in making the form available for optical scanning data entry. The issue of serving size is also reviewed, including whether to ignore serving sizes and simply ask about frequency, to ask specific questions about serving size, or to specify a typical serving size. A major advantage of standardized food frequency questionnaires is that they can be self-administered and optically scanned. This has made large prospective studies feasible, in particular repeated assessments of diets. Available evidence suggests that, in a reasonably literate population, mailed self-administered questionnaires and detailed interviewer-administered questionnaires provide similar degrees of validity.

Reproducibility and validity of food frequency questionnaires

Although practical considerations greatly favour the use of food frequency questionnaires in epidemiologic studies, it is also crucial to consider in detail the degree to which such questionnaires measure true dietary intake. Reproducibility of a questionnaire is an important, but not sufficient, indicator of the value of dietary data obtained by that method. In general, validity, which involves a comparison with a superior method, will be a more important criterion. In assessing the validity of dietary assessment methods, an important consideration is whether errors in the comparison method are correlated with errors in the method under evaluation. The use of biochemical measures to assess validity has appeal in that errors are likely to be highly independent. However, they do not provide a quantitative evaluation of validity and, for many nutrients, no practical biochemical measure exists. Thus, most validation studies have used primarily either dietary record data or 24-hour-recall data for comparisons. Dietary record data have the advantage that the cognitive processes are quite different than those for food frequency questionnaires; thus, the errors are less likely to be correlated.

A large number of validation studies have now been performed in a variety of populations comparing food frequency questionnaire information with diet

records or 24-hour recalls. In general, with comprehensive questionnaires, correlations of 0.6–0.7 appear to be attainable for nutrient intakes adjusted for total energy. Although this degree of validity is not perfect, it will be enough to assure that important associations will not be missed if the study size is sufficiently large. In this section, a variety of alternatives for expressing validity are described.

For this topic, a computer problem is usually included in which students are asked to analyse data from a validation study and present the results. Subsets of data for three or four nutrients from the comparison of a food frequency questionnaire with repeated diet records are used for this purpose (Willett 1985; Rimm 1992). To discourage use of student work from previous years, we extract different subsets of data from year to year.

Biochemical indicators of dietary intake

This section focuses on the principles underlying the use of biochemical indicators in nutritional epidemiology. A comprehensive review of available biochemical indicators is not attempted. A fundamental distinction is made between the use of a biochemical measurement as an indicator of dietary intake (which is the primary rationale in nutritional epidemiology) *versus* being of inherent interest as a predictor of disease. Serum cholesterol is used as an example that is a poor indicator of dietary intake but is still of interest in predicting risk of coronary heart disease. Important considerations in the evaluation of a potential biochemical indicator are whether it is sensitive to intake. Many biochemical measurements such as serum calcium or sodium are tightly regulated and thus represent diet poorly. The ability of a biochemical measure to integrate intake over an extended period of time is also critical. It is also important to identify other non-dietary factors that influence the level of biochemical indicators; if they are known, it may be possible to adjust for these factors to eliminate them as extraneous sources of variation.

The use of biochemical indicators in epidemiologic studies is also considered. Important issues are the timing of sampling, and proper processing and storage to avoid artefacts or degradation. The use of biochemical indicators in nested case-control studies is also covered, with particular attention to the avoidance of bias. In general, the usefulness of biochemical indicators of diet in epidemiologic studies can be viewed as a spectrum: at one end, dietary methods perform poorly and only biochemical measurements will be useful (for example, selenium intake); in the middle, both dietary and biochemical measurements may be similarly useful (e.g. beta-carotene and folic acid); and at the other end, only dietary intake data is likely to be useful (such as calcium or dietary fibre intake).

Particular attention is given to the monitoring of laboratory precision in epidemiologic studies, as this can be a major cause of error.

The evaluation of biochemical indicators of diet for use in epidemiologic studies is also described. Approaches include the evaluation of repeated measurements in the same individuals over time. If there is not a reasonably high correlation, a biochemical indicator will not provide useful information on long-term intake for an individual. Whether the measure is sensitive to dietary intake can be evaluated by comparisons between intake and biochemical levels in cross-sectional studies or by the manipulation of dietary intake in intervention studies. Typically, we use several examples to illustrate the use of biochemical indicators for epidemiologic applications.

For the topic of biochemical indicators of diet, students are required to analyse a dataset relating dietary intakes to blood levels of several nutrients. Specific attention is given to the effects of adding covariates on the primary association. As an example, we have used data comparing intakes of carotenoids, retinol, and vitamin E assessed by food frequency questionnaires with their corresponding blood levels (Willett 1983) but any similar dataset would suffice.

Anthropometric measures and body composition

Anthropometric variables, particularly weight and height, are the most commonly employed measures of nutritional status in epidemiologic studies, mainly because of their simplicity and ease of collection. Measures of body dimensions and mass can be used directly but most often combinations of weight and height have been used to estimate the relative body composition, such as fatness. Compared to other biologic measurements, height and weight can be assessed with great precision in epidemiologic studies, even by self-report within most populations. The limitation of body mass index as a measure of adiposity relates to the failure of weight to distinguish between fat mass and lean mass. In middle-aged general populations in western countries, it does appear that the large majority of variation of body weight-adjusted-for-height is due to differences in fat mass. However, this assumption holds up less well in other groups, such as body builders, the elderly, and quite probably persons in developing countries, where physical labour is more common.

Several new methods have become available for measuring body composition, including electrical impedance and dilution methods. However, the superiority of these compared to standard anthropometric measurements has yet to be documented for epidemiologic applications. Many studies have addressed the relative value of various combinations of height and weight to

assess body fatness. The use of external criteria such as biochemical measurements sensitive to body fatness is described as a method of evaluating relative validity. In general, the use of the simple measure of weight divided by height squared (body mass index) appears to be at least as useful as other combinations and has the virtue of being widely used and comparable across studies.

Recently, great interest has emerged in the use of measurements to assess the distribution of body fat, with the assumption that intra-abdominal fat has metabolic properties that are distinct. The use of circumference measurements alone or in combination to assess body fat distribution is discussed. Similarly, the value of skin folds is considered, although these are generally unlikely to be practical in most epidemiologic investigations. The virtue of body circumference measurements is that they can be assessed by individuals themselves with relatively good validity. As an example, the controversies regarding the setting of weight guidelines are reviewed (Willett et al. 1999; Hu et al. 2004; Flegal et al. 2005; Zhang et al. 2008).

Implications of total energy intake for epidemiologic analyses

Total energy intake deserves special consideration in nutritional epidemiology for several reasons. First, energy intake may be a primary determinant of disease. Second, individual differences in total energy intake produce variation in intake of specific nutrients because the consumption of most nutrients is positively correlated with total energy intake; this added variation may be extraneous to disease risk and thus a source of error. Third, when energy intake is associated with disease but not a direct cause, the effects of specific nutrients may be confounded by total energy intake.

The topic of total energy intake is introduced by a discussion of energy physiology. Notably, total energy intake is largely determined by lean body mass and level of physical activity. Thus, unless a person is willing to change physical activity or body fat substantially, total energy intake for an individual is relatively fixed. Thus, to alter intake of specific nutrients, it is necessary for an individual to do this primarily via a change in the composition of their diet rather than via a change in total food intake. This is fundamentally important in nutritional epidemiology because the implication is that we need to study primarily the relation of dietary composition, rather than absolute nutrient intake, to disease risk.

A number of methods to account or adjust for total nutrient intake in nutritional epidemiologic studies are described. The classical approach has been to use nutrient density (nutrient divided by calories). The inherent problem with

this is that the nutrient density will remain confounded if total caloric intake is a predictor of disease. Other approaches include adding total energy intake as a covariant or utilizing the residual method to isolate variability in nutrient intake independent of total energy intake. An energy partition model can also be used but this does not control for energy intake. Illustrations are given of situations where inadequate or inappropriate adjustment for energy intake can actually reverse the direction of associations between dietary factors and disease risk. Thus, careful attention to total energy intake is essential in nutritional epidemiologic analysis. For readings, we use chapter 11 of *Nutritional Epidemiology* and Willett (1997).

Correction for measurement error

Forms of measurement error in epidemiologic studies are considered. The most common assumption is that measurement error for an individual is simply random within-person variation, as usually assumed for 24-hour recalls. Systematic errors can occur within individuals or groups, particularly when assessing diet using structured methods such as food frequency questionnaires. The effects of these various types of errors on associations are considered in detail.

Until recently, it has usually been simply noted that measurement error will tend to bias associations to the null. However, it is possible to measure the degree of error and utilize correction methods to obtain an estimate of association that would exist if there had been no measurement error. In our introductory course several simple methods for correctional methods are provided. These include the correction of standard deviations for distributions of a single variable, correction of correlation coefficients, and correction of relative risks using the regression calibration approach. The effect of measurement errors in covariates and correction of these errors is also described. In addition, issues regarding the use of imperfect 'true' measurements as the standard, and the potential implications of correlated measurement errors between the standard and surrogate method, are described in principle. The topic of correction for measurement error is a rapidly expanding area in epidemiology, and the current literature exceeds what can be included in an introductory course.

As part of this section, students are usually given a computer problem set involving the correction of standard deviations, the de-attenuation of correlation coefficients, and the correction of relative risks in a univariate context.

Analysis and presentation of dietary data

The analysis and presentation of data from epidemiologic studies is not inherently different from other aspects of epidemiology, but some topics are particularly

important because of the complexity of nutritional data. A common issue is the handling of missing information on foods and outlying values. Several approaches are discussed. Advantages and disadvantages of using nutritional data in categorical or continuous forms are also addressed. Alternatives in the graphic presentation of the data are described with illustrations and discussions of advantages and disadvantages.

Temporal relationships in the study of most chronic diseases can be critical. Most frequently, we do not know the temporal relationship between diet and the diagnosis of disease, although at times we have reasons to hypothesize specific induction periods. An exploration of alternative relationships can provide information that can be inherently valuable. In addition, when no association is observed, this conclusion will be most compelling if a wide range of induction periods is evaluated. A variety of analytic strategies are discussed.

Multivariate approaches to the data analysis can be particularly important because dietary factors tend to be correlated, sometimes strongly. However, a number of serious pitfalls can exist in many situations. The inclusion of the covariate can radically alter the biologic meaning of the original variable even though the analysis might be logical from a purely statistical standpoint. In addition, the inclusion of multiple variables can dramatically reduce the remaining variation, leading to an analysis that is essentially uninformative. Another common situation arises when one or more variables are components of another, such as saturated, monounsaturated, and polyunsaturated fat, which are all components of total fat. Various options for analytic strategies are discussed that address somewhat different questions.

Interest in overall dietary patterns has generated various approaches to creating dietary scores. These factors of methods are discussed. Other topics include the analysis and interpretation of subgroups and interactions.

The use of meta-analysis for systematically summarizing results of randomized trials has become routine, but its place in epidemiology has been controversial. Meta-analysis of published data from dietary studies is particularly problematic because of the wide variety of ways that data have been analysed and presented. The use of pooled analyses based on primary data is particularly attractive because this will allow a consistent analytic approach in studies (Smith-Warner et al. 2006).

Substantive examples of nutritional epidemiology applications

A variety of topics are now available that can be used to illustrate and reinforce an understanding of issues in nutritional epidemiology. Although applications and examples are helpful when interwoven into issues discussed throughout

the course, a focused discussion of a specific topic can be valuable. Specific topics that have been used in our course follow.

Vitamin A and lung cancer

This topic has a long history in nutritional epidemiology, beginning with the early report of an apparent protective effect of vitamin A against lung cancer. Interest shifted to the hypotheses that beta-carotene might be a responsible factor for the remarkably consistent inverse association between food and vegetable consumption, and risk of lung cancer. However, randomized trials of high-dose beta-carotene supplements did not support any evidence of benefit and suggested possible harm (The Alpha-Tocopherol Beta-Carotene Cancer Prevention Study Group 1994). This area illustrates many issues, including the use of biochemical indicators, potential pitfalls of extrapolating from data on foods to a specific nutrient, biases in case-control studies, and the strengths and limitations of randomized trials for testing hypotheses.

Dietary fat and breast cancer

The relation between dietary fat and breast cancer has been a major focus in nutritional epidemiology because prospective cohort studies have generally not supported the hypotheses based on ecologic correlations. The possible explanations for these findings are considered, including lack of validity of dietary assessment, an inadequate range of dietary intake, inappropriate specification of the induction period, and a true lack of association. Various means of addressing these alternative explanations are described. The limitations of randomized trials, in which an effect of dietary fat reduction on breast cancer incidence has also not been supported, are also discussed.

Diet and coronary heart disease

This has been one of the longest-standing issues in public health for the last fifty years. The original diet–heart hypothesis focused on saturated fat and dietary cholesterol and became a widespread belief but with limited empirical support. More recent evidence suggests that this hypothesis was highly incomplete as it did not take into account the strong benefits of unsaturated fatty acids and the adverse effects of *trans* fatty acids. Moreover, recent evidence has suggested that many other dietary constituents can have important impacts on coronary heart disease risk; these factors include the characteristics of dietary carbohydrate, omega-3 fatty acids, folic acid and vitamin B_6, dietary antioxidants, and other micronutrients. The topic illustrates the importance of multivariate analysis in nutritional epidemiology and the integration of metabolic studies, observational epidemiology, and randomized trials.

Folic acid and neural tube defects

The relationship between folic acid intake and risk of neural tube defects has prompted a major paradigm shift in nutrition. This most clearly illustrates the enormous impact of differences in micronutrient intake without accompanying signs of clinical deficiency. The topic is a rich source of teaching material and includes contributions of randomized trials, case-control and cohort studies using dietary intake data, and biomarkers. Recent data also illustrate the use of genetic polymorphisms in nutritional epidemiology and how they may greatly enhance the causal interpretation of associations. The topic also illustrates the application of nutritional epidemiologic approaches after a causal association has been established, because the existence of an effect raises questions about optimal intake and public health approaches to achieve optimal levels (Mason et al. 2007).

Advanced topics in nutritional epidemiology

The basic material in an introductory course in nutritional epidemiology is now too substantial to allow in-depth examination of current methodologic issues in this field. Moreover, for many students, such an in-depth discussion is neither necessary nor appropriate. Thus, a course on advanced topics can be valuable for those who will be actively engaged in research.

An advanced topics course would appropriately include readings and discussions on current literature in the field of nutritional epidemiology, particularly focusing on methodologic issues. A large part of this literature presently involves issues regarding measurement error and correction for measurement. In our department, Drs Alberto Ascherio and Donna Spiegelman have developed computer-simulation problems based on actual data from our large cohort studies and validation sub-studies. These problems allow students to investigate the impacts of measurement error and to use correction procedures. In addition, the advanced course can allow students to analyse diet and disease relationships using actual data.

Supervised research

As in any doctoral programme, much of the actual learning process occurs in the context of supervised original research. An ideal student experience would include not just the analysis of existing data but also participation in a full cycle of hypothesis development, grant writing, data collection, data cleaning and processing, analysis of data, and reporting of research findings in written and oral form. The complete process from beginning to end will almost always take

more than the usual number of years for completion of a doctoral degree. Thus, typically we attempt to have students involved in all aspects of this work but not necessarily with all aspects directly related to the same topic.

Assessing students' achievements

For our basic course in nutritional epidemiology, student achievement is assessed by the following:

- written reports based on homework problems;
- completion of their own dietary assessments, including the analysis of diet records using a standard food composition analysis system;
- classroom discussion, which encourages completing the assigned readings; and
- a final examination covering key concepts in the course.

Conclusion

In a relatively short period, nutritional epidemiology has developed into a major area of activity within the overall field of epidemiology. Nutritional epidemiology has generated methodologic work that has implications broadly across the overall field. Because of the complexity of nutritional epidemiology, training in this field will provide students with concepts and skills that can be applied in many other aspects of epidemiology. The coming years are sure to provide a major growth in available data because many large prospective dietary studies are only now beginning to provide results. These will provide an enormous expansion in knowledge of diet and health and should hopefully stimulate further the refinement and development of epidemiologic methods.

References

Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group (ATBC). (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine*, **330**: 1029–35.

Baer, H. J., Colditz, G. A., Rosner, B., Michels, K. B., Rich-Edwards, J. W., Hunter, D. J., and Willett, W. C. (2005) Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: a prospective cohort study. *Breast Cancer Research*, **7**: R314–25.

Beaton, G. H., Milner, J., Corey, P., McGuire, V., Cousins, M., Stewart, E., de Ramos, M., Hewitt, D., Grambsch, P. V., Kassim, N., and Little, J. A. (1979) Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *American Journal of Clinical Nutrition*, **32**: 2546–9.

Eliassen, H. A., Colditz, G. A., Rosner, B., Willett, W. C., and Hankinson, S. E. (2006) Adult weight change and risk of postmenopausal breast cancer. *Journal of the American Medical Association*, **296**: 193–201.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. (2005) Excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association*, **293**: 1861–7.

Heady, J. A. (1961) Diets of bank clerks: development of a method of classifying the diets of individuals for use in epidemiologic studies. *Journal of the Royal Statistical Society*, **124**: 336–61.

Hennekens, C. H. et al. (1996) Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine*, **334**: 1145–9.

Hu, F. et al. (1997) Dietary fat intake and the risk of coronary heart disease in women. *New England Journal of Medicine*, **337**: 1491–9.

Hu, F. B., Willett, W. C., Li, T., Stampfer, M. J., Colditz, G. A., and Manson, J. E. (2004)Adiposity as compared with physical activity in predicting mortality among women. *New England Journal of Medicine*, **351**: 2694–703.

Kim, E. H., Willett, W. C., Colditz, G. A., Hankinson, S. E., Stampfer, M. J., Hunter, D. J., Rosner, B., and Holmes, M. D. (2006) Dietary fat and risk of postmenopausal breast cancer in a 20-year follow-up. *American Journal of Epidemiology*, **164**: 990–7.

Linos, E., Willett, W. C., Cho, E., Colditz, G., and Frazier, L. A. (2008) Red meat consumption during adolescence and risk of breast cancer among premenopausal women. *Cancer Epidemiology Biomarkers and Prevention*, **17**: 2146–51.

Mannisto, S. et al. (2004) Dietary carotenoids and risk of lung cancer in a pooled analysis of seven cohort studies. *Cancer Epidemiology Biomarkers and Prevention*, **13**: 40–8.

Margetts, B. and Nelson, M., eds. (1997) *Design Concepts in Nutritional Epidemiology* (2nd edn). Oxford: Oxford University Press.

Mason, J. B., Dickstein, A., Jacques, P. F., Haggarty, P., Selhub, J., Dallal, G., and Rosenberg, I. (2007) A temporal association between folic acid fortification and an increase in colorectal cancer rates may be illuminating important biological principles: a hypothesis. *Cancer Epidemiology Biomarkers and Prevention*, **16**: 1325–9.

Michels, K. B. and Xue, F. (2006) Role of birthweight in the etiology of breast cancer. *International Journal of Cancer*, **119**: 2007–25.

Rimm, E. B., Giovannucci, E. L., Stampfer, M. J., Colditz, G. A., Litin, L. B., and Willett, W. C. (1992) Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *American Journal of Epidemiology*, **135**: 1114–26.

Shekelle, R. B. et al. (1981) Diet, serum cholesterol, and death from coronary heart disease: the Western Electric Study. *New England Journal of Medicine*, **304**: 65–70.

Smith-Warner, S. A. et al. (2006) Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. *American Journal of Epidemiology*, **163**: 1053–64.

Willett, W. C. (2013) *Nutritional Epidemiology* (3rd edn). New York: Oxford University Press.

Willett, W. C., Dietz, W. H., and Colditz, G. A. (1999) Guidelines for healthy weight. *New England Journal of Medicine*, **341**: 427–34.

Willett, W. C., Howe, G. R., and Kushi, L. H. (1997) Adjustment for total energy intake in epidemiological studies. *American Journal of Clinical Nutrition*, **65 Suppl. 4**: 1220S–8S.

Willett, W. C., Sampson, L., Stampfer, M. J., Rosner, B., Bain, C., Witschi, J., Hennekens, C. H., and Speizer, F. E. (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology*, **122**: 51–65.

Willett, W. C., Stampfer, M. J., Underwood, B. A., Speizer, F. E., Rosner, B., and Hennekens, C. H. (1983) Validation of a dietary questionnaire with plasma carotenoid and alpha-tocopherol levels. *American Journal of Clinical Nutrition*, **38**: 631–9.

Zhang, C., Rexrode, K. M., van Dam, R. M., Li, T. Y., and Hu, F. B. (2008) Abdominal obesity and the risk of all-cause, cardiovascular, and cancer mortality: sixteen years of follow-up in US women. *Circulation*, **117**: 1658–67.

Chapter 12

Genetic epidemiology

Harry Campbell and Susan Service

Introduction to genetic epidemiology

A major challenge for epidemiology in the future is understanding the role of genetic risk factors and how genetic factors interact with environmental factors in causing disease. There is a clear need for all epidemiologists to understand how sequencing of the human genome and developments in genetic technology have increased the potential for investigation of genetic risk factors. Genetic epidemiology concerns itself with questions such as the following:

- What is the prevalence of gene variants in different populations?
- What is the risk (absolute and relative) of disease associated with these variants?
- What is the contribution of gene variants to the occurrence of disease in different populations (i.e. attributable risk)?
- What is the risk of disease associated with gene–gene and gene–environment interactions?
- What is the validity of genetic tests?

To read relevant textbooks and follow leading articles published in the field of genetic epidemiology, it is important that students have or acquire a minimum literacy in the use of basic genetics terms and concepts and have some knowledge of basic genetic laboratory techniques. A number of 'primers' have been published in textbooks or appear on the Internet, and teachers should recommend one they can make easily accessible to their students (see the bibliography). There are relatively few books that could serve as a course text in genetic epidemiology.

Genetic technology advances so rapidly that new approaches to the study of genetic risk factors become possible and quickly supersede previous approaches that become less favoured or obsolete, making it difficult if not impossible for a textbook to be up to date. It is more likely, therefore, that background reading to complement teaching will be based on selected recent journal articles or web resources rather than on one or two textbooks. It is important that students are directed towards and introduced to relevant journals and websites in which the

latest methods are described and discussed. This serves as an adjunct to coursework but also supports students' subsequent education in genetic epidemiology after they have completed the formal coursework. Journals that fit into this category (such as *American Journal of Human Genetics*, *PLoS Genetics*, *Nature Genetics*, and *Nature Reviews Genetics*) can be found in the bibliography. In addition, general epidemiology journals such as *American Journal of Epidemiology*, *Epidemiology Reviews*, *Epidemiology*, and *International Journal of Epidemiology* are increasingly publishing articles of investigations of genetic risk factors, and reviews of genetic epidemiology study designs. These are usually rather more comprehensible to the student than the more detailed discussion found in specialist journals such as *Genetic Epidemiology*.

Teaching objectives

An introductory course should concentrate on principles and important concepts, together with a general introduction to specific methods or analytic procedures. More detailed exploration of specific methods is best targeted at students who are currently undertaking research which requires the application of these methods since, as noted above, the precise details of methods and analysis programmes are regularly evolving and improving. A short course for undergraduate students could cover basic genetic concepts, define genetic epidemiology and the type of research questions it addresses, give a concise description of the main approaches, and note their uses, advantages, and disadvantages. A postgraduate course could follow the curricular outline given below. This could be presented as a module within a general MSc programme in epidemiology, a stand-alone course, or as a complete MSc in genetic epidemiology. The length of the course would govern the depth of coverage and amount of 'hands-on' practical sessions working with computer programs. Longer courses would normally contain a supervised assignment or research project that requires students to apply the knowledge and skills acquired in the course. All courses should equip students with a basic understanding of how to address the following hierarchy of research questions in genetic epidemiology:

- Does the disease cluster in families?
- Is the clustering caused by genetic or environmental factors?
- Is there evidence for a genetic factor? (Can a specific mode of inheritance be identified?)
- Can the genetic risk factor be localized and identified as causal?
- Do environmental factors modify the expression of the genetic factor? (Is there evidence of gene–environment interaction?)

Teaching content

The exact content of the course will depend on the students' background, their particular interests, and the length of time available for the course. One suggestion for the structure of an introductory course in genetic epidemiology for those interested in understanding common complex disorders (rather than rare Mendelian disorders) is given in Box 12.1.

Box 12.1 Structure of an introductory course in genetic epidemiology

1. Genetics review
   a. Basic principles of genetics
   b. Laboratory techniques (optional)
2. Genotypes and phenotypes
   a. Hardy–Weinberg equilibrium
   b. Penetrance
   c. Heterogeneity
   d. Modes of inheritance (dominant, recessive, additive/multiplicative)
   e. Quantitative vs qualitative traits
   f. Endophenotypes
   g. Phenomics
   h. Population genetics concepts (genetic drift, founder effects, coalescent theory, mutation, selection)
   i. Other concepts (anticipation, imprinting)
3. How to study if a trait has a 'genetic' component
   a. Twin/adoption/admixture/migration studies
   b. Segregation analysis
   c. Measurement of the risk ratio, lambda
   d. Using genetic similarity between individuals to make inferences about heritability
   e. Characteristics of traits that are amenable to genetic mapping
   f. Characteristics of populations that favour the study of genetic factors

Box 12.1 Structure of an introductory course in genetic epidemiology *(continued)*

4. How to identify whether specific genetic factors are related to a trait
   a. Approaches to complex traits
      i. Genome screens (linkage and genome-wide association)
      ii. Fine mapping
      iii. Candidate gene studies
      iv. Whole-genome and whole-exome sequencing data
      v. The use of gene expression data to complement mapping results
      vi. The need for international collaborative, multidisciplinary approaches
   b. Important epidemiologic considerations
      i. Ascertainment bias
      ii. Sample size
      iii. Multiple testing
      iv. Confounding
      v. Causal inference
5. Major study designs
   a. Family-based designs
   b. Population designs based on linkage disequilibrium
      i. Traditional epidemiologic approaches (evaluating association between a trait and genotypes/alleles using odds ratios)
      ii. Genome-wide association studies
      iii. Methods that incorporate population genetics (coalescent approaches, shared segment analyses)
      iv. Linkage disequilibrium across the genome and implications for genetic mapping of common, complex traits: the HapMap project
   c. Strengths and weaknesses of each approach
6. How to study gene–environment and gene–gene interactions
   a. Study designs
   b. Sample size
   c. Quantification of risk

Box 12.1 Structure of an introductory course in genetic epidemiology *(continued)*

d. Exploratory data-mining approaches
e. Computational issues

7. Using genetic data to study environmental exposures (Mendelian randomization studies)
8. Laboratory genetic analysis
   a. DNA collection and storage
   b. Genotyping methods and arrays; genome sequencing data
   c. Genotype misclassification/sequencing errors
9. Functional studies of possible causative variants
10. Ethical and public health issues
    a. The impact of genetic research on patients and their families
    b. Issues in genetic testing
11. Computer programs for data management and data analysis

Teaching method and format

Principles and concepts can be taught by didactic methods but are best illustrated with many examples from the literature and from the teacher's own experience. Many concepts may be difficult to grasp initially. The liberal use of examples reinforces the points made by approaching the concept from a different angle and may aid understanding by setting what may seem as abstract concepts into a specific context. It can also help pace the teaching so that the course does not simply present a large number of new ideas without a break and thus risk students becoming lost or losing concentration. Examples can be presented in a way that invites active participation by the class and is therefore a useful way of checking whether or not students have understood new concepts.

It is essential that there is a strong 'hands-on' element, with computer practicals so that students can gain experience in handling data and can be introduced to computer programs used for analysis. At the very least this should be provided as a demonstration, with output from the analysis provided to students and used in exercises or homework. Ideally, students should have the opportunity for guided practice. However, there are relatively few people using 'genetic epidemiology' programs, and methods are constantly being updated. This has resulted in there being less investment in making user-friendly interfaces to

these programs than in other areas of epidemiology. Furthermore, different areas of genetic epidemiology (e.g. linkage analysis as compared to association analysis) may use entirely different software programs and operating systems for data analysis; this can be very confusing for students and presents a problem for teaching computer practical sessions. In an introductory course much time can be wasted instructing students in the mechanics of how to use a specific program, and command-driven programs with user-unfriendly interfaces can be particularly difficult for students. Teachers should avoid devoting any significant time to this unless the course specifically aims to teach skills in a particular technique or with specific programs. When the computer programs are complicated to operate, it is better either to give a demonstration or to give each student a detailed set of instructions of the steps required so time is not wasted. Many students may have had experience only in point-and-click style computer operating systems. In designing the practical computer exercises for an introductory class, particularly with undergraduates, it might be best to use as few different programs as possible and, when given a choice of program, use those that will operate with a GUI interface in a desktop environment. Even when it may not be possible for the students to actually run an analysis themselves, it can be very instructive to examine the raw output from such an analysis that the teacher has performed. Learning to interpret and understand the output can be more valuable than learning to 'push the buttons' to run an analysis.

Teachers should end each session by summarizing the aims of the session and reviewing what was done. They should also highlight general principles that can be taken from the example. This helps to offset the tendency for students to become so involved in details of the operation of the program that they lose sight of the relevance of what they are doing. These sessions should also give details of where these programs can be obtained. An excellent example of a web resource is the GenABEL project (see 'Analysis software') which aims to provide 'a framework for collaborative open-source based development of statistical genomics methodology' and includes software, manuals, tutorials, and access to peer support. When calculations can be performed easily by use of a simple spreadsheet, opportunities should be given for students to tackle the analysis of illustrative datasets. Examples of where this may be appropriate include calculating expected genotype counts under Hardy–Weinberg equilibrium or using data to illustrate the effect of population stratification on risk estimates in association studies.

Once the principles of a particular method have been taught and an example given in a computer demonstration or a guided worked example, then it may be helpful to give students a reprint of a recent publication which uses this approach and ask them to read this critically overnight. This can then be discussed as a

class exercise the following day. This is a useful way of reinforcing understanding of the method by setting it in context. It can also be used to check that students have grasped the key issues and for introducing further refinements or more advanced issues for individual follow-up by students. The ability to critically read and interpret scientific literature is an important skill for students to develop. Another useful exercise is to have students (working in teams) present to the class a journal article that uses approaches/techniques discussed in lecture. As students may not be very familiar with reading and dissecting scientific literature, it may be helpful to provide them with generic questions (e.g. what is the research hypothesis, what is the target population, what are the inclusion/exclusion criteria, what are the analysis methods, etc.) to guide and focus their analysis and presentation. Encourage students to display figures and tables from the articles (either digitally or on an overhead projector) with the expectation that they will be able to explain and interpret them.

In some courses the teacher may wish to consider setting aside time to demonstrate a few basic techniques in a genetics laboratory. This investment of time is, depending on the background of students, worthwhile as part of the introductory section of a longer course which is run over several weeks or months. An appreciation of the complexity and validity of laboratory methods can help students read published reports critically and interpret results appropriately.

Teaching notes

This section will follow the curricular outline described above, will draw attention to some key issues to cover, and will identify teaching resources which could be used to develop teaching materials on these topics. There are many textbooks and web resources which can be used to explain basic genetic concepts (see examples in the list of websites in the bibliography). It is essential that the teacher check understanding of concepts as they occur in the course or, if necessary, at the outset of the course.

The relationship between genotypes and phenotypes

Genotypes are composed of individual alleles, which may or may not act independently in determining the probability of expressing a phenotype. It is worth spending time ensuring that students understand the Hardy–Weinberg equilibrium and its implications in the relationship between genotype and phenotype. Students should be shown how to count allele frequencies and how to calculate genotype frequencies in the next generation when given known allele frequencies. They should be aware of some of the applications of measuring or comparing allele frequencies (e.g. in comparing two populations or in association studies).

Students should understand the concept of incomplete penetrance and conditional probability—that the probability of expressing a phenotype is dependent upon the genotype.

The study of endophenotypes is a powerful means by which to discover genetic factors that underlie risk for complex disease traits. These are measures that are related to the dichotomous trait of interest but are often less subjectively measured and can be evaluated in persons without the full trait phenotype (for example serum measurements of IgE in a study of asthma). Many endophenotypes are quantitative, enabling the use of powerful quantitative trait linkage methods. The field of phenomics seeks to understand the interrelationship of endophenotypes and classic disease categorizations and evaluate whether endophenotypes are heritable and therefore possible to genetically map.

Population genetics concepts

Concepts that are relevant to an understanding of genetic epidemiological study designs or data analysis should be mentioned and students given references for further reading on these topics. Thus effective population size (N_e), genetic drift, and founder effects may be relevant to the choice of study population in genetic association and linkage studies. An understanding of coalescent theory can help to understand some approaches to the analysis of genetic association data. Knowledge of the effects of natural selection can help in the interpretation of data (e.g. the marked geographical variation of the *LCT* gene (lactose tolerance)). Similarly an understanding of the concepts of anticipation and imprinting may be important in the correct interpretation of study findings. As sequence-level datasets become more common, a firm footing in the basics of population genetics becomes important to interpret the frequency and distribution of rare variation.

How to study whether a trait has a 'genetic' component

The role of more traditional epidemiologic study designs such as adoption studies, admixture studies, and migration studies has been reviewed in the textbook by Khoury et al. (1993). The teacher may wish to review briefly the role of the more traditional approaches in estimating the relative contribution of genetic and environmental components of disease risk.

Twin studies can measure the contribution of a genetic component to total variation: this is based on measuring concordance between twins. Students should understand that concordance can be quoted in two different ways (pairwise or proband-wise). They should be familiar with the correct interpretation of the main comparisons between monozygous and dizygous twins (reared together or apart) and their other siblings. They should understand that

the calculated heritability estimates are not absolute but may vary in different environmental settings. A published example of estimating heritability by means of a twin study should be discussed with students.

The risk ratio, lambda, is based on the measurement of the (increased) risk of disease for relatives of cases compared to the population risk. This can be calculated for all types of relative pairs but is most often quoted for siblings of cases (lambda sib). This serves as a measure of familial clustering. Since this clustering could be in whole or in part caused by environmental factors, an adoption study would be required to determine the relative importance of genetic and environmental sharing. Lambda is often estimated in studies that recruit cases retrospectively from clinic records and then study disease in relatives. It should be highlighted that this tends to lead to overestimation of lambda, since families with multiple cases tend to be selected. There are also problems with determining the affection status of unaffected relatives of young age in adult onset disease and with diseases that influence reproduction. Comparison of values of lambda for different sets of relatives can give clues as to the nature of the underlying genetic model (as has been done for schizophrenia) but is highly dependent on the precision of estimates published in the literature.

Estimating a parameter such as lambda sib or carrying out segregation analysis can be an important first step in studying the role of genetic factors in a disease. It can provide direction concerning whether or not further investment of resources into trying to identify specific alleles is justified and, if so, which approach may be more appropriate. It is a critical determinant of the power of affected sib pair studies.

Segregation analysis investigates whether familial clustering is consistent with Mendelian transmission patterns by estimating the proportion of affected offspring and comparing this to the expected proportion based on Mendelian inheritance. An introduction to segregation analysis should begin with an explanation of types of ascertainment (complete, single incomplete, and multiple incomplete), ascertainment bias, and the methods of correction for this bias in the different types of ascertainment. Students should be made aware of the general principles of segregation analysis and where to obtain computer programs for this analysis. However, computer practicals should be left to more advanced courses for those who intend to use this approach in their work.

For quantitative phenotypes, evaluating whether a quantitative trait is significantly heritable can establish a genetic component to the phenotype. Students should understand and work an example of the estimation of heritability from the parent-offspring regression, and be introduced to the estimation of heritability from pedigrees. With the availability of dense single nucleotide polymorphism (SNP) genotype data, it is also possible to obtain rough heritability

estimates from population, rather than family, collections, by comparing phenotypic similarity with genome-wide genetic similarity, and students should be introduced to these methods as well.

How to identify whether specific genetic factors are related to a phenotype and major study designs

Approaches to complex traits

Until the advent of genetic maps in the 1980s, identifying the gene responsible for a phenotype relied heavily on knowledge of the biological basis of the genetic defect itself, a strategy termed 'functional cloning'. While functional cloning has been successful for a few diseases, including phenylketonuria, sickle cell anaemia, and haemophilia A, for the vast majority of single gene disorders, no such biological information exists. As more anonymous genetic polymorphisms (markers) began to be identified and located on the human genetic map, the position of a gene likely responsible for a disorder was inferred by the co-segregation of the phenotype and the genetic markers, a process called genetic linkage. 'Positional cloning' is the term coined to describe the identification of a gene by first localizing its position using genetic markers, with no knowledge whatsoever of the gene product. Function of the gene is determined only after it has been identified. This strategy is often referred to as a genome screen. Subsequent 'fine mapping' usually involves genotyping a dense set of markers in the region of interest. It may be that the function of some of the genes within the implicated region is known and, from their function and position, they can be considered good candidates as disease loci. Discussion of the history of gene mapping serves to demonstrate to the students how advances in technology have shaped the way genetic epidemiology is conducted.

Students should understand the differences between the strategies of genome screening, fine mapping, and candidate gene analysis, and understand when each strategy is preferred. Genome screening, fine mapping, and candidate gene studies can be performed on both quantitative and qualitative phenotypes. Both family-based designs and population-based designs can be used for genome screening. Fine mapping is usually performed in population samples but may be done in family-based designs. Candidate gene studies are nearly always done in population samples. The different study designs and analysis methods are discussed below.

Pedigree studies and (parametric) linkage analysis

This approach is particularly suited to studying genes with major effects on disease risk. It is one method of mapping disease susceptibility genes and involves analysing segregation patterns in families with multiple affected members.

Introductory teaching should start by covering concepts of recombination, recombination fraction, genetic distance, and, if necessary, likelihood theory (maximum likelihood estimates and likelihood ratio test). Properties of the LOD (logarithm of the likelihood ratio in favour of linkage) score should be outlined: that LOD scores greater than 0 generally favour linkage, that LOD scores can be added across families (at a common value of the recombination fraction), and that the maximum LOD score over all recombination fractions less than 0.5 provides a good test statistic for the presence of linkage. The interpretation of maximum LOD scores should be discussed by outlining the theoretical and empirical basis for the traditional position of taking a maximum LOD score of greater than 3 as evidence of linkage. The influence of multiple tests such as that which occurs in genome-wide screening should also be discussed. Students should understand that, since the LOD score is a 'conditional likelihood of the markers given the disease', linkage analysis is robust to ascertainment bias and hence is valid to select high-risk families for study.

Similarly, although linkage analysis is based on an underlying model, the type 1 error rate (significance level) is valid even when the model parameters are mis-specified. However, mis-specification of marker allele frequencies (and in particular assuming an allele is rarer than it actually is) can invalidate LOD score results and lead to false-positive findings. The relative advantages of two contrasting approaches to specifying allele frequencies, that is, using published frequencies from large samples from the same ethnic group or using maximum likelihood estimates from observed allele counts over all study pedigrees, can be compared.

Simple example pedigrees (assuming that the Mendelian inheritance pattern is known and that the marker and disease status of everyone is known and assuming single genes, no phenocopies, and full penetrance) can be used to hand-count the number of recombinant and non-recombinant offspring and calculate a LOD score. Examples with increased complexity (unknown disease status) can then be introduced and discussed and the concept of phase should be illustrated through these examples. Example tables of LOD scores at various recombination fractions and LOD score curves should be given to the class for group discussion of their interpretation. The use of computer simulation methods to assess the power of a family or set of families to detect linkage should be described.

Once basic principles have been covered, concepts such as incomplete penetrance (penetrance functions and liability classes), inbreeding loops, and genetic heterogeneity (heterogeneity LOD score) can be introduced. The mechanics of how the LOD score is calculated in more complex settings should be presented. The importance of considering genetic heterogeneity in circumstances in which

more than one gene may be involved in the etiology of the disease under study should be emphasized. The teacher should describe the strategy of calculating posterior probabilities of linkage for each family (once linkage to the first locus is established) to identify families from which to search for linkage with other markers (fine mapping). Finally, the principles of multipoint analysis can be presented. Fine mapping is dependent on correct model specification, and the results of multipoint analysis are therefore very sensitive to model misspecification (quantitative trait locus (QTL); methods which are discussed below tend to be more stable). Examples of the successful use of this approach (for example, in breast cancer) should be used as a way of reviewing and summarizing this approach. Students should be introduced to linkage programs by means of a demonstration or closely guided worked example. Details should be given of how to access available linkage analysis programs and related textbooks (see the bibliography). If students are to be taken beyond this point to gain competence in handling the individual programs, then there are a number of computer-based courses run regularly which are designed for this purpose (see 'Short courses in genetic epidemiology').

Pedigree studies and (non-parametric) linkage analysis

This approach restricts attention to those affected and assesses whether there is more sharing of alleles which are identical by descent (IBD) in groups of affected relatives than you would expect if there were no linkage. An introduction to this method should cover the limitations of parametric linkage analysis when studying complex disease. The teacher should check that students understand the concept of IBD. More general non-parametric tests aimed at detecting linkage by testing for increased marker similarity between the affected members of a pedigree can be explained in outline. The problem of distinguishing between IBD and identical by state (IBS) in adult onset disease (when parents may not be available for study), and the use of other siblings in an attempt to reconstruct parental genotypes or of adjacent markers to provide additional information to help distinguish IBD and IBS should be discussed. Students should be asked to calculate means tests or chi-square values for various patterns of observed sharing and then to derive values of lambda as a way of reinforcing their understanding of underlying principles.

A number of issues pertinent to both parametric and non-parametric linkage analysis should be reviewed at this stage. A review of linkage statistics is helpful: LOD scores, location scores (in multipoint linkage programs), and non-parametric linkage scores, since these have different interpretations, with a LOD score of 3 being roughly equivalent to a location score of 13.8 or a non-parametric linkage score of 3.7. The calculation of 95 per cent confidence limits

or 1 or 3 LOD limits should be described, and the fact that stringent LOD score criteria are particularly important for genome searches should be highlighted.

Pedigree studies and quantitative trait linkage analysis

Genetic epidemiologic approaches for the study of continuous traits in family-based samples should be described. Students should understand why looking at a population distribution is not a good way of identifying major genetic effects. The potential gain of information from this approach compared to the analysis of qualitative traits should be noted but with the recognition that there have been fewer examples of successful mapping of QTL in humans than in animal models. In part, this may be because of insufficient power. Just as the value of lambda is a critical determinant of the power in non-parametric linkage analysis, so the value of rho (the proportion of variance caused by the gene under study) is critical for the power of QTL analysis. To have sufficient power, 10–15 per cent of the variance observed should be caused by the genetic locus under study. Limiting the variance from environmental factors (for example, by studying smaller family structures such as twins) might be one way of increasing this value.

Association studies (allelic association) in population samples

The principles of this approach are similar to that of a traditional case-control study with their inherent strengths and weaknesses. The approach is based on showing a higher or lower allele frequency among cases than controls. In most studies the (marker) allele under investigation is a marker which was very close to the disease-causing allele at the time in history when this mutation arose and has remained in linkage disequilibrium with the disease-causing allele since that time. However, this approach can also be used to study candidate genes presumed to include the disease-causing alleles. It should be pointed out that this approach studies association of alleles (not genetic loci), unlike linkage analysis, which studies the linkage of genetic loci (not alleles).

Since there are some differences in approach to traditional case-control studies, it is important that students are taken through the steps in an association study, starting with checking that both the control and case alleles are in Hardy–Weinberg equilibrium. Lack of Hardy–Weinberg equilibrium may be due to chance, laboratory typing errors, or population stratification and can be explored by checking laboratory results and studying other markers. Only when Hardy–Weinberg equilibrium holds can allele counts be assumed to be independent and thus the comparison of allele frequencies be considered valid. If either set of alleles are not found to be in Hardy–Weinberg equilibrium, then a comparison of genotype counts (combining genotypes with small numbers, if

necessary) is more valid but has considerably less power. Alternatively, one can assume that the alleles act additively (multiplicatively, in the case of a qualitative phenotype) and use the Cochran–Armitage test for trend. Simple datasets from case-control studies can be given to students to hand-tally results and so gain experience in the analysis of data using all three tests (alleles, genotypes, and the trend test). The teacher should check for understanding of the concept of linkage disequilibrium and revise this, if necessary. Fundamentally, in case-control samples, analysis and interpretation of genetic risk factors is not hugely different from the analysis and interpretation of other risk factors for disease, and the teacher should draw these parallels, as students coming from a strictly epidemiological background may feel more comfortable in this setting. For qualitative phenotypes, students should understand the interpretation of the odds ratio for allelic, genotypic, and trend tests. For quantitative phenotypes, basic concepts of ANOVA and regression should be presented and the interpretation of regression coefficients discussed.

Students should appreciate the problem of interpreting positive results in association tests and be aware of the many examples in the literature of reported positive associations which have not been repeated by subsequent studies. Strategies for distinguishing between the various explanations for observed positive associations—chance, population stratification, linkage disequilibrium, and causal association—should be presented. These include repeating the study in a different population with the same or different study design, investigating evidence of association with several markers in different areas of the genome (to detect stratification effects), and studying markers close to those reported to show association (since association to these markers might be expected if linkage disequilibrium were the likely explanation). The latter should be checked in the control and the case chromosomes, since background linkage disequilibrium can be fairly extensive in some isolate populations. It is also possible that a true association may not be replicated in another study, and possible explanations for this outcome should be discussed, including, for example, low power in the replication sample and genetic heterogeneity. Students should understand alternative methods to identify and control for population stratification in case-control samples such as genomic control, correction for population stratification using eigen-analysis, and use of IBS data and linear mixed models to correct for both population stratification and cryptic relatedness. The pros and cons of the different approaches should be discussed.

Sample size considerations should be described and a comparison made between the power of population-based approaches compared to parametric and non-parametric linkage analysis of complex diseases, as mentioned above. Meta-analysis of comparable datasets from several studies within research

consortia is a common way of achieving much larger sample sizes, and students should understand the principles of this technique (see Evangelou and Ioannidis 2013).

Students should understand the basic design and analysis approaches for genome-wide association studies. This should include an understanding of the available genotyping platforms and genotype arrays, algorithms employed to call genotypes, batching of samples and related quality control issues (e.g. use of Hardy–Weinberg equilibrium checks and quantile–-quantile (QQ) plots). It is important to realize that laboratory errors in genotyping occur. Differential bias in genotype scoring between cases and controls can occur when samples are handled differently and scored separately and this can result in test statistic inflation and false positives. Details of design features such as multistage approaches, the role of genetic enrichment strategies in the first stage (such as including cases with early onset or positive family history), and methods of selecting which SNPs are taken forward to next stage should be covered. Similarly there should be discussion of analysis issues including statistical approaches to testing of single SNP associations and methods of measuring or correcting for population stratification (such as the use of QQ plots of Cochran–Armitage test statistics with correction for any inflation factor). The relative power of joint versus separate analysis of data from different stages and need for replication samples should be covered. More general issues of data sharing and public posting of complete databases (timing, problems, utilization) should be discussed. Students should be introduced to tools that can be used to manage the very large datafiles generated by genome-wide association studies (e.g. PLINK, which contains tools for data management, summary statistics, population stratification, association analysis, and identity-by-descent estimation (see resources section). In addition, there are several available genetic analysis courses which cover the basic principles and methods of genome-wide association studies (see 'Short courses in genetic epidemiology').

Once both this linkage and association analyses have been covered in the course, the advantages and disadvantages of these two approaches for the study of complex disease should be compared. One dimension of this should be to discuss their relative power to detect genetic risk factors of modest to low effect.

Approaches using the coalescent and analysis of shared segments

There is an increasing interest in the potential of the analysis of shared segments in isolate populations. This is based on the identification of shared IBD segments which have come from a common ancestor. The larger the number of generations since the common ancestor, the smaller is the expected size of the

shared segment. Approaches such as homozygosity mapping and ancestral haplotype reconstruction work best in populations in which there is a high probability that patients share a recent common ancestor. Of course, at some point in the past, all individuals will share some set of common ancestors, even in non-isolate populations. Coalescent theory guides the modelling of when the gene genealogy of any two individuals comes together in a common ancestor, and has been adapted to mapping disease genes. The principles underlying these methods should be outlined. A discussion of the favourable characteristics of populations to support these studies can serve as a useful session to introduce or review understanding of concepts from population genetics such as founder effects, genetic drift, coalescent theory, and genetic heterogeneity. Many of these methods are computer intensive and may not serve well as exercises, except for very simple examples.

Linkage disequilibrium and the International HapMap Project

The HapMap study was an international collaboration to genotype SNPs at a very high density (one SNP every 2–5 kb) through the entire genome in 270 persons from four major population groups (European, African, Japanese, and Chinese). As the project evolved, samples from additional populations were genotyped. All the data from this project are available for download. These data present a wonderful tool for the examination of the extent and distribution of linkage disequilibrium in different genomic regions and populations. The HapMap project and its resources should be presented and their use in the design and execution of association studies discussed. The students should understand how varying patterns of linkage disequilibrium have implications for the optimal SNP density to detect association.

Genome sequencing

Exome and whole-genome sequencing are becoming widely available and affordable in genetic research, and a number of specific courses are now available to provide training in handling and analysing these data (see 'Short courses in genetic epidemiology'). Next-generation sequencing methodologies (Shendure and Ji 2008) have reduced the cost of sequencing very dramatically. Both whole-genome and whole-exome approaches (Kiezun et al. 2012) are now widely employed. The depth to which genomes are sequenced has great impact on the quality of the resulting data, and the degree to which rare variants can be identified (Sims et al. 2014). Interpretation of results relies heavily on bioinformatics and prediction of function of missense variants. Rare variants lack power to be analysed individually for association, so a host of aggregation

methods to examine the impact of the 'burden' of rare variants on phenotype have been developed.

The 1000 Genomes Project has extensive sequence data available for several populations. If one or more of these populations provide a suitable reference for the population under study, then imputation methods can be used in conjunction with genotype data to obtain in silico sequence data. Students could review imputation methods and their assumptions and discuss the trade-off between imputation and sequencing (e.g. expense vs ability to discover unique variation; Li et al. 2009).

Copy number variation

Genomic variability encompasses not only SNPs and short tandem repeats but also structural alterations such as deletions, duplications, and inversions. In particular, copy number variants (CNVs) are now known to account for a substantial amount of genetic variation. Increasingly, inherited CNVs are being linked to human disease (e.g. by influencing gene expression) and somatic chromosomal alterations involved in cancer. Students should understand that different technologies which can generate CNV data and their strengths and weaknesses in identifying CNVs. International efforts are being made to define appropriate nomenclature and create a database of known human structural variations.

Gene expression

Genetical genomics has been proposed to map loci controlling gene expression differences (eQTLs) that might underlie disease trait variation. Correlation of gene expression with phenotype levels, in conjunction with genetic mapping/association results (for both phenotype and expression), can be used to identify genes involved in phenotype variability. Students should understand the difference between *cis* and *trans* regulation of gene expression, and the basics of how gene expression data are generated (chip expression vs RNA-seq expression).

Epigenetic epidemiology

Epigenetic variation can be genetically or environmentally determined and can contribute to variation in gene expression among individuals and thus cause variation in common complex disease risk (Relton and Davey Smith 2010). There are courses available which present the principles and methods of epigenetic epidemiology (see 'Short courses in genetic epidemiology').

Gene–environment and gene–gene interactions

A major challenge in the study of the role of genetic factors in complex disease is that causal pathways may involve a number of genetic variants interacting

with each other and with environmental factors. It is likely that individual genetic variants will have a small effect at an individual level (although at a population level they may be important—as measured by population-attributable risk). It is helpful to review the various possible mechanisms of gene–environment interaction and how these can be identified. Ways of adapting the various approaches already covered in the course to study gene–environment interaction should be described. The power of various study designs for detecting gene–environment or gene–gene interaction should be discussed with students. The instructor should point out that power to detect interactions, for a given sample size in a case-control study, is usually much weaker than the power to detect genetic main effects, and that examination of all possible gene–environment or gene–gene interactions may add a substantial computational burden in a genome screen setting. The 'case-only' design has increased power to detect interactions with other genes and the environment over a case-control design. The instructor should explain that the case-only design will detect only departures from the multiplicative model of risk, and in order to apply it one must assume independence between exposure to the environmental risk factor and genotype in the population. Similarly, in order to use this design to study gene–gene interactions, one must assume that the frequencies of the genes are independent in the population (see Yang and Khoury 1997 for more information on the case-only design).

Multivariate data-mining approaches may aid in the detection of unknown relationships among genes or between genes and environmental factors. Data mining involves the analysis of large observational datasets to find unsuspected relationships and to summarize the data in new ways that are understandable and useful. Data mining has in the last decade emerged as a strong area of research at the intersection of computer science and statistics. The presentation of these approaches should include a discussion of how one might control for spurious findings that can come from data 'fishing expeditions'.

Mendelian randomization studies

The theory underpinning the Mendelian randomization (MR) approach is based on the random assortment of alleles transferred from parent to offspring at the time of gamete formation. This random assortment of alleles at conception is equivalent to a randomized controlled trial in which people are randomly allocated to different genotypes rather than therapeutic interventions and in which a functional genetic variant acts as a proxy for an environmental exposure. The main concept of an MR study is based on three relationships: genotype–intermediate phenotype, intermediate phenotype–disease, and genotype–disease (Didelez and Sheehan 2007; Lawlor et al. 2007). It can be used to

identify causal environmental risk factors without the several potential problems of observational epidemiology (Smith and Ebrahim 2004). In particular, if population stratification is controlled, the distributions of genetic variants are generally independent of confounding factors (e.g. behavioural, social, physiological, or environmental; Didelez and Sheehan 2007; Lawlor et al. 2007). The MR approach can also strengthen causal conclusions by limiting reverse causation problems (biological, through exposure assignment, due to reporting bias), selection bias, and regression dilution bias (Smith and Ebrahim 2004). The power of this approach has been increased through the use of meta-analysis of multiple datasets to increase sample size for analysis.

Functional studies

While identifying genetic variants associated with a phenotype is usually a long, complex process in and of itself, the teacher should emphasize that this is just one step in understanding how a gene may control expression of a phenotype. The teacher should point out that positive linkage or association findings do not, in and of themselves, establish causality in any way. Fine mapping techniques such as resequencing the locus which contains the variant showing association may be required to identify further genetic variation in the study (or other populations) for further study. It will be important to consider biological data on function together with epidemiological data. An example of these steps is described in detail in the article by Lowe et al. (2007). The instructor should remind students of the differences between synonymous and non-synonymous mutations, and between missense and nonsense mutations, and emphasize that it is rare that studies will discover a missense or nonsense mutation that obviously has a clear alteration of protein function. It is more likely to be the case, especially when studying complex disorders, that the variant(s) may instead affect risk of disease by more subtle mechanisms, such as altering gene expression. It is quite difficult to separate benign variation from possible functional variation, and there is no simple road map or recipe on the best way to demonstrate a functional role for a variant (see Prokunina and Alarcon-Riquelme 2004 for a review). The strategy will usually involve some combination of in silico (bioinformatic predictions) and in vitro (laboratory experiments) work. Examples of in vitro functional studies include splicing studies to evaluate if an allele is affecting how a gene is spliced, and reporter studies to investigate allele-specific regulatory effects on the expression of a reporter gene. The teacher should point out possible problems in such studies related to differences in tissue type used in the experiment (e.g. only lymphocytes may be available but the expression of the gene more related to the phenotype under study would only be seen in brain tissue). Furthermore, environmental interactions, and differences

in tissue quality and preparation between individuals may complicate interpretation of results. In silico studies include algorithms to predict transcription binding sites, and comparison of conserved sequences across species.

Ultimately it is in vivo studies that will demonstrate the causal role of a gene/variant in the expression of a phenotype. Specific mutations can be inserted into experimental animals (usually mice), or the gene itself can be 'knocked out' (made non-functional). The instructor should explain that not all human phenotypes are amenable to animal models, and it may take creative and careful work to define an animal phenotype that is parallel to the human phenotype. Environmental interactions should not be ignored in this stage, as even when investigating a true genetic risk factor, phenotypes may not be expressed in animal models without an environmental contribution. After a causal link between a specific variant and a phenotype has been established, it will still be necessary to screen this gene in other persons with the same phenotype to identify additional causal mutations – for example the *BRCA1* gene that increases risk for breast cancer has thousands of known mutations. It is possible that newly ascertained breast cancer patients from families with strong risk for breast cancer may harbour an entirely new mutation in this gene.

Ethical and public health issues

Establishing a causal link between a gene/mutation and phenotype enables persons at risk for the phenotype to be genetically tested for mutations. The ethical and social implications of testing subjects for known causal mutations should be discussed with students. Whole-genome or whole-exome sequencing data gives the possibility to identify deleterious variants not related to the phenotype under study (e.g. identifying breast cancer mutations in persons ascertained for a study of lipid disorders) and this can raise ethical issues involved with relaying such information to study subjects.

The teacher should explain the concepts of sensitivity, specificity, and the positive predictive value of a genetic test. Design issues in genetic risk prediction studies should be mentioned (Janssens et al. 2011). Students should discuss what value a genetic test outcome might have for a subject, if any, when there is no additional treatment for the phenotype that could be applied as a result of knowing the subject's mutational status.

Computer programs for data management and data analysis

The online resources list URLs where many computer programs commonly used in the analysis of genetic data can be found. The programs vary a great deal in their ease of use, the operating system they use, and the type of analyses they will perform. No one package will perform all types of analyses possible in

genetic epidemiology. Several websites now contain suites of statistical genetics software, some of which is linked to collaborative networks that can provide peer support (see 'Websites'). The instructor should emphasize that, as this is a growing and changing field, analysis methods and software evolve very quickly, and one must be flexible enough to adapt to these changes. The instructor should also mention that, in real applications, studies will involve thousands of persons and hundreds of thousands of markers. Efficient data management programs are needed to manipulate this volume of data and get it into the correct format(s) for analysis. Laboratories involved in genetic epidemiology often employ not only statisticians to analyse the data, but also data management and IT specialists to handle these large datasets. Guidelines for good practice in computational research should be followed (Sandve et al. 2013).

The Human Genetic Epidemiology Network (HuGENet)

Students should be introduced to HuGENet: a global collaboration of individuals and organizations committed to the assessment of the impact of human genome variation on population health and how genetic information can be used to improve health and prevent disease (<http://www.cdc.gov/genomics/hugenet/default.htm>). HuGENet has defined standards for reporting results from genetic association studies, for replicating genome-wide association studies, and for assessing the quality of evidence in favour of an association. Web resources aim to assemble knowledge from genetic epidemiology studies in various formats. Noteworthy resources include HuGE Navigator (<http://www.hugenavigator.net/>), which provides access to a continuously updated knowledge base including information on population prevalence of genetic variants, gene–disease associations, gene–gene and gene–environment interactions, and evaluation of genetic tests and allows users to search integrated information on genetic associations, published HuGE literature, HuGE investigators, candidate genes, association data extracted from meta-analysis publications and HuGE field synopses generated by domain experts, as well as analysis of temporal and geographic trends.

Students should also be made aware that extensive datasets are being made available for reference and individual level data available for analysis through resources such as dbGaP (**d**ata**b**ase of **G**enotype **a**nd **P**henotype), which was developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype (see 'Datasets'). These include genome-wide association studies, medical sequencing, and molecular diagnostic assays, as well as association studies between genotype and non-clinical traits. dbSNP is an online resource that has details of all identified genetic variation which can support research in a wide range of areas including physical mapping and population genetics (see 'Datasets').

Support after the course

After an initial course, students can be informed about specific courses that focus solely on practical instruction in specific techniques. These are run by a number of different agencies, including those listed in the web resources. Students should also be encouraged to read the journals noted in 'Introduction to genetic epidemiology' regularly as new methods are usually published first in these journals (e.g. see current issues of relevant journals listed in the bibliography). Inform the students of any ongoing seminar series or journal clubs at the university and encourage their attendance and participation.

Assessing students' achievements

Short courses of undergraduate teaching in genetic epidemiology should check students' understanding of important concepts and principles. A variety of methods, including short-essay or multiple-choice questions, can be adopted depending on the overall assessment strategy for the course. At the postgraduate level the summative assessment should, in addition, address the application of these principles and the handling and analysis of simple datasets. This is probably best achieved by a supervised research project. This project can be supplemented by assessments requiring critical appraisal of published studies and analysis of simple datasets (using either spreadsheet or specialist genetic epidemiology computer programs which were used in the course). Teachers may also choose to include an assessment of the students' understanding of the public health, ethical, social, or legal implications of advances in the understanding of the role of genetic factors in common diseases. Examples of opportunities for checking understanding and giving feedback through formative (in course) assessment have been given throughout this chapter.

Bibliography

Original articles

General

Barton, N. H. and Keightley, P. D. (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**: 11–21.

Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nature Genetics*, **33**: 228–37.

Freimer, N. and Sabatti, C. (2003) The human phenome project. *Nature Genetics*, **34**: 15–21.

Lander, E. S. and Schork, N. J. (1994) Genetic dissection of complex traits. *Science*, **265**: 2037–48.

Wright, A., Charlesworth, B.,Rudan, I., Carothers, A., and Campbell, H. (2003) A polygenic basis for late-onset disease. *Trends in Genetics*, **19**: 97–106.

Family-based designs and methodology

Benyamin, B., Visscher, P. M., and McRae, A. F. (2009) Family-based genome-wide association studies. *Pharmacogenomics*, **10**: 181–90.

Kruglyak, L., Daly, M. J., Reeve-Daly, M. P., and Lander, E. S. (1996) Parametric and nonparametric linkage analysis: a unified approach. *American Journal of Human Genetics*, **58**: 1347–63.

Laird, N. M. and Lange, C. (2006) Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, **7**: 385–94.

Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, **11**: 241–7.

Population-based designs and methodology

Cardon, L. R. and Bell, J. I. (2001) Association study designs for complex diseases. *Nature Reviews Genetics*, **2**: 91–9.

de Koning, D. J. and Haley, C. S. (2005) Genetical genomics in humans and model organisms. *Trends in Genetics*, **21**: 377–81.

Do R., Kathiresan, S., and Abecasis, G. R. (2012) Exome sequencing and complex disease: practical aspects of rare variant association studies. *Human Molecular Genetics*, **21**: R1–9.

Evangelou, E. and Ioannidis, J. P. (2013) Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, **14**: 379–89.

Goldstein, D. B., Allen, A., Keebler, J., Margulies, E. H., Petrou, S., Petrovski, S., and Sunyaev, S. (2013) Sequencing studies in human genetics: design and interpretation. *Nature Reviews Genetics*, **14**: 460–70.

Hirschhorn, J. N. and Daly, M. J. (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, **6**: 95–108.

Kiezun, A. et al. (2012) Exome sequencing and the genetic basis of complex traits. *Nature Genetics*, **44**: 623–30.

Li Y., Willer, C., Sanna, S., and Abecasis, G. (2009) Genotype imputation. *Annual Review of Genomics and Human Genetics*, **10**: 387–406.

Little, J. et al. (2009) Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *European Journal of Epidemiology*, **24**: 37–55.

Manolio, T. A. et al. (2009) Finding the missing heritability of complex diseases. *Nature*, **461**: 747–53.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, **9**: 356–69.

Relton, C. L. and Davey Smith, G. (2010) Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Medicine*, **7**: e1000356.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, **15**: 121–32.

The International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, **437**: 1299–320.

Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012) Five years of GWAS discovery. *American Journal of Human Genetics*, **90**: 7–24.

Wang, W. Y., Barratt, B. J., Clayton, D. G., and Todd, J. A. (2005) Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics*, **6**: 109–18.

Zondervan, K. T. and Cardon, L. R. (2004) The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*, **5**: 89–100.

Gene–gene and gene–environment interaction

Hunter, D. J. (2005) Gene-environment interactions in human diseases. *Nature Reviews Genetics*, **6**: 287–98.

Marchini, J., Donnelly, P., and Cardon, L. (2005) Genome wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics*, **37**: 413–17.

Yang, Q. and Khoury, M. J. (1997) Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiologic Reviews*, **19**: 33–43.

MR studies

Didelez, V. and Sheehan, N. (2007) Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, **16**: 309–30.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., and Davey Smith, G. (2007) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, **27**, 1133–63.

Smith, G. D. and Ebrahim, S. (2004) Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology*, **33**: 30–42.

Quantitative traits

MacKay, T. F. C. (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics*, **35**: 303–39.

Members of the Complex Trait Consortium. (2003) The nature and identification of quantitative trait loci: a community's view. *Nature Reviews Genetics*, **4**: 911–16.

Zhang, H. and Risch, N. (1996) Mapping quantitative-trait loci in humans by use of extreme concordant sib pairs: selected sampling by parental phenotypes. *American Journal of Human Genetics*, **59**: 951–7.

Population genetics and coalescent theory

Rosenberg, N. A. and Nordborg, M. (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**: 380–90.

Admixture mapping

Smith, M. W. and O'Brien, S. J. (2005) Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nature Reviews Genetics*, **6**: 623–32.

Laboratory genetic analysis

Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**: 847–59.

Shendure, J. and Ji, H. (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**: 1135–45.

Functional studies of possible causative variants/animal studies

Buckland, P. R. (2006) The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochimica et Biophysica Acta*, **1762**: 17–28.

Flint, J., Valdar, W., Shifman, S., and Mott, R. (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nature Reviews Genetics*, **6**: 271–86.

Lowe, C. E. et al. (2007) Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nature Genetics*, **39**: 1074–82.

Peters, L. L., Robledo, R. F., Bult, C. J., Churchill, G. A., Paigen, B. J., and Svenson, K. L. (2007) The mouse as a model for human biology: a resource guide for complex trait analysis. *Nature Reviews Genetics*, **8**: 58–69.

Prokunina, L. and Alarcon-Riquelme, M. E. (2004) Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Reviews in Molecular Medicine*, **6**: 1–15.

Ethical and public health issues

Janssens, A. C., Ioannidis, J. P., van Duijn, C. M., Little, J., and Khoury, M. J. for the G. R.I. P.S Group. (2011) Strengthening the reporting of Genetic Risk Prediction Studies: the GRIPS statement. *Genetics in Medicine*, **13**: 453–6.

Janssens, A. C. J. W., Pardo, M. C., Steyerberg, E. W., and van Duijn, C. M. (2004) Revisiting the clinical validity of multiplex genetic testing in complex diseases. *American Journal of Human Genetics*, **74**: 585–8.

The International HapMap Consortium. (2004) Integrating ethics and science in the International HapMap Project. *Nature Reviews Genetics*, **5**: 467–75.

Computer programs for data management and data analysis

Balding, D. J. (2006) A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, **7**: 781–91.

Excoffier, L. and Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**: 745–58.

Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. (2013) Ten simple rules for reproducible computational research. *PLoS Computational Biology*, **9**: e1003285.

Examples of applications to specific diseases

Bell, C. G., Walley, A. J., and Froguel, P. (2005) The genetics of human obesity. *Nature Reviews Genetics*, **6**: 221–34.

Cowley, A. W. Jr. (2006) The genetic dissection of essential hypertension. *Nature Reviews Genetics*, **7**: 829–40.

Farrer, M. J. (2006) Genetics of Parkinson disease: paradigm shifts and future prospects. *Nature Reviews Genetics*, **7**: 306–18.

Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J., and Krawczak, M. (2005) Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nature Reviews Genetics*, **6**: 376–88.

Watkins, H. and Farrall, M. (2006) Genetic susceptibility to coronary artery disease: from promise to progress. *Nature Reviews Genetics*, **7**: 163–73.

Books

Balding, D. J., Bishop, M., and Cannings, C., eds. (2001) *Handbook of Statistical Genetics*. Chichester: John Wiley & Sons.

Ewens, W. J. and Grant, G. R. (2001) *Statistical Methods in Bioinformatics*. New York: Springer-Verlag.

Falconer, D. S. and Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics* (4th edn). London: Longman.

Hartl, D. L. and Clark, A. G. (1997) *Principles of Population Genetics* (3rd edn). Sunderland, MA: Sinauer Associates, Inc.

Khoury, M. J., Beaty, T. H., and Cohen, B. H. (1993) *Fundamentals of Genetic Epidemiology*. Oxford: Oxford University Press.

Khoury, M. J., Bedrosian, S., Gwinn, M., Higgins, J. T., Ioannidis, J. P.A., and Little, J. (2010) *Human Genome Epidemiology* (2nd edn). New York: Oxford University Press.

Lange, K. (1997) *Mathematical and Statistical Methods for Genetic Analysis*. New York: Springer-Verlag.

Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc.

Neale, M. C. and Cardon, L. R. (1992) *Methodology for Genetic Studies of Twins and Families*. Dordrecht: Kluwer Academic Publishers.

Ott, J. (1991) *Analysis of Human Genetic Linkage* (revised edn). Baltimore, MA: The John Hopkins University Press.

Sham, P. (1998) *Statistics in Human Genetics*. New York: John Wiley & Sons.

Strachan, T. and Read, A. P. (1996) *Human Molecular Genetics*. Oxford: BIOS Scientific Publishers.

Thomas, D. (2004) *Statistical Methods in Genetic Epidemiology*. Oxford: Oxford University Press.

Weiss, K. M. (1993) *Genetic Variation and Human Disease: Principles and Evolutionary Approaches*. Cambridge: Cambridge University Press.

Websites

General resource gateways for genetic epidemiology/genomics research

Centers for Disease Control and Prevention. (2014) *Public Health Genomics*. <http://www.cdc.gov/genomics/>, accessed 6 November 2014. This site provides updated information on how human genomic discoveries can be used to improve health and prevent disease. It also provides links to US Centers for Disease Control activities in public health genomics; of particular relevance is the Human Genetic Epidemiology Network (HuGENet; <http://www.cdc.gov/genomics/hugenet/>). HuGENet is a global collaboration of individuals and organizations committed to the development and dissemination of population-based information on the human genome; the website acts as an information exchange network on genetic epidemiology and contains links to resources such as the HuGE Navigator (<http://www.hugenavigator.net/>), which is a human genome epidemiology knowledge base that can be searched in a variety of ways (e.g. via Genopedia, which searches genetic associations and human epidemiology summaries on the basis of gene names (<http://www.hugenavigator.net/HuGENavigator/startPagePedia.do>).

Details of selected recent articles on genetic epidemiology can be found at <http://www.cdc.gov/genomics/update/current.htm#SciLit>.

Garlipp, K. and Zollmann, F. S. (2005) *HUM-MOLGEN: Current Issues of Biomedical Journals*. <http://hum-molgen.org/journals/>, accessed 6 November 2014. This website contains links to relevant online genetics journals.

PHG Foundation. (2014) *PHG Foundation*. <http://www.phgfoundation.org>, accessed 6 November 2014. The PHG (Public Health Genetics) Foundation is a charitable organization dedicated to the translation of biomedical sciences and in particular of genome-based knowledge and technologies, in order to benefit the health of individuals and populations. The website contains a regular newsletter with commentary on recently published important articles on genetic epidemiology or genetics and public health, and details of a wide range of relevant web resources, including those in genetic epidemiology, genetics education and training, genetics dictionaries and glossaries, links to relevant online journals, and ethical, legal, and social implications of human genetics.

US Department of Energy Human Genome Project. (2013) *Human Genome Project Information*. <http://web.ornl.gov/sci/techresources/Human_Genome/project/index.shtml>, accessed 6 November 2014. This website contains an enormous variety of genetics resources, including a primer on molecular genetics (<http://web.ornl.gov/sci/techresources/Human_Genome/education/index.shtml>), a directory of education resources in genetics, a glossary of genetic terms, and links to relevant online genetics journals.

Datasets

International HapMap Project. (2013) *The International HapMap Project*. <http://www.hapmap.org>, accessed 6 November 2014. This project has genotyped 270 persons of European, Asian, or African ancestry for over 6 million single nucleotide polymorphisms (SNPs). Patterns of association among these SNPs have been analysed and all data are available for download. The data are a wonderful resource for example datasets and are used extensively by researching selecting a parsimonious set of SNPs for trait mapping.

National Center for Biotechnology Information, U.S. National Library of Medicine. (2014) *dbGaP*. <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap>, accessed 6 November 2014. The dbGaP (a database of Genotype and Phenotype) was developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype. These studies include genome-wide association studies, medical sequencing, and molecular diagnostic assays, as well as association studies between genotype and non-clinical traits.

National Center for Biotechnology Information, U.S. National Library of Medicine. (2014) ***dbSNP***. <http://www.ncbi.nlm.nih.gov/snp>, accessed 6 November 2014. The dbSNP (a database of Single Nucleotide Polymorphisms) is a free public archive of genetic variation within and across different species; the archive was developed and is currently hosted by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI).

Analysis software

Broad Institute. (2014) *Haploview*. <http://www.broad.mit.edu/mpg/haploview/>, accessed 6 November 2014. Haploview is a free programme that expedites the analysis of haplotype data. The programme is easy to use and interfaces nicely with HapMap data.

GenABEL Project. (2014) *GenABEL.org.*<http://www.genabel.org/>, accessed 6 November 2014. Free open-source software and collaborative development of statistical genomics methodology.

Laboratory of Statistical Genetics, The Rockefeller University. (2014) *Computer Programs*. <http://lab.rockefeller.edu/ott/programmes>, accessed 6 November 2014. This website contains a suite of freely downloadable programmes, including LINKAGE, FASTLINK, and GENEHUNTER, that can be used for parametric and non-parametric analysis of pedigree data.

Montana, G. (2014) *CRAN Task View: Statistical Genetics*. < http://cran.r-project.org/web/views/Genetics.html>, accessed 6 November 2014. R is a free downloadable programme for interactive statistical analysis, and several researchers have contributed add-on packages for use in R. The link above is to the genetics package. Other packages of possible interest can be found at <http://cran.r-project.org/web/views/>.

Purcell, S. (2014) *PLINK . . . Whole Genome Association Analysis Toolset*. <http://pngu.mgh.harvard.edu/purcell/plink/>, accessed 6 November 2014. PLINK is an open-source C/C++ whole genome association analysis tool set. Large datasets comprising hundreds of thousands of markers genotyped for thousands of individuals can be rapidly manipulated and analysed using PLINK. It has five main domains of function: data management, summary statistics, population stratification, association analysis, and identity-by-descent estimation. Further details are given in Purcell S et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis, *American Journal of Human Genetics*, **81**: 559–75.

Purcell, S. and Sham, P. (2009) *Genetic Power Calculator*. <http://pngu.mgh.harvard.edu/~purcell/gpc/>, accessed 6 November 2014. This interactive online calculator provides an easy way to estimate the power of a sample in order to detect association to a trait. Several different designs, including case-control and family-based control, are available.

SAS Institute Inc. (2014) *SAS/Genetics(TM) 9.2 User's Guide*. <http://support.sas.com/documentation/cdl/en/geneug/59659/HTML/default/viewer.htm#geneug_intro_sect001.htm>, accessed 6 November 2014. SAS is a commonly used statistical analysis package. SAS is not free but most universities have site licences for SAS, and many epidemiology students might be familiar with SAS from other work. The SAS Genetics package could be a useful add-on for those individuals already working in this environment.

UCLA Human Genetics. (2014) *UCLA Human Genetics Software Distribution*. <http://www.genetics.ucla.edu/software/>, accessed 6 November 2014. This site hosts Mendel, a comprehensive package for exact statistical genetic analysis of qualitative and quantitative traits and which is freely available for download. Mendel can analyse both association and pedigree data under a variety of models. Numerous example files are provided that could be used as teaching tools.

University of Michigan Center for Statistical Genetics. (2014) *Gonçalo's Software*. <http://www.sph.umich.edu/csg/abecasis/software.html>, accessed 6 November 2014. Gonçalo Abecasis leads a small research group in the University of Michigan's Department of Biostatistics' Center for Statistical Genetics. His group is focused on developing the computational and statistical tools required for understanding human genetic variation, with a particular focus on complex human disease. He has a variety of useful, free programmes for download, including Merlin for linkage analysis.

Yang, J. (2013) *GCTA: a Tool for Genome-Wide Complex Trait Analysis*. <http://www.complextraitgenomics.com/software/gcta/>, accessed 6 November 2014. Software to estimate the proportion of phenotypic variance explained by genome analysis and many other types of analyses, to better understand the genetic architecture of complex traits.

Teaching resources

Centers for Disease Control and Prevention. (2014) *Public Health Genomics*: *Genomic Testing*. <http://www.cdc.gov/genomics/gtesting/ACCE/acce_proj.htm>, accessed 6 November 2014. ACCE, which takes its name from the four components of evaluation—analytic validity, clinical validity, clinical utility, and associated ethical, legal and social implications—is a model process for evaluating data on emerging genetic tests. A table on the webpage lists a series of targeted questions aimed at a comprehensive review of genetic testing and which could be interesting to discuss with students.

www.oege.org. (2010) *OEGE–Online Encyclopedia for Genetic Epidemiology Studies*. <http://www.genes.org.uk/software/>, accessed 6 November 2014. This website provides a variety of useful links for genetic epidemiology, including links to online calculators for the Hardy–Weinberg equilibrium and other statistical functions.

Short courses in genetic epidemiology

Laboratory of Statistical Genetics, The Rockefeller University. (2014) *List of Short Courses of Genetic Analysis*. <http://lab.rockefeller.edu/ott/shortcourses>, accessed 6 November 2014. This website provides a listing of a number of short courses on genetic analysis.

Netherlands Institute of Health Sciences. (2014) *NIHES Research Training in Medicine and the Health Sciences*. <http://www.nihes.nl>, accessed 6 November 2014. NIHES offers a number of short courses, approximately one week in duration, that cover many topics in genetic epidemiology.

UCLA Human Genetics. (2014) *Courses*. <http://www.genetics.ucla.edu/courses/>, accessed 6 November 2014. The Department of Human Genetics at the University of California, Los Angelos, usually offers annually a short course in statistical genetics. Much of the work in the course focuses on the use of the software package Mendel. However, general principles and lessons can be carried over to other settings.

Wellcome Trust. (2014) *Advanced Courses and Scientific Conferences*. <http://www.wellcome.ac.uk/Education-resources/Courses-and-conferences/Advanced-Courses-and-Scientific-Conferences/>, accessed 6 November 2014. The list of courses provided by this website includes ones on human genome analysis: genetic analysis of multifactorial diseases and design, and analysis of genetic-based association studies.

Chapter 13

Teaching molecular epidemiology

Betsy Foxman

Introduction to teaching molecular epidemiology

Molecular epidemiology has been defined in many ways (Foxman and Riley 2001); differences between definitions are more attributable to a lack of clarity on what constitutes 'molecular' than coming from the definition of 'epidemiology. 'Molecular' has become a synonym for modern molecular techniques that characterize nucleic and amino acids and sometimes also includes metabolites (the 'omics': genomics, transcriptomics, proteomics, and metabolomics). Even this definition of 'molecular' is narrow, as it excludes many laboratory techniques applied to biologic material that might be usefully included in the study of the distribution and determinants of population health and disease. I prefer to define 'molecular' as any laboratory technique applied to biological material. This preference reflects both the ongoing, rapid development in laboratory methods, and my experience that, regardless of the actual technique, the underlying principles of integrating laboratory techniques with epidemiology are the same. Thus, I focus my course on the implications of the merger of 'molecular'—however defined—with the design, conduct, and analysis of epidemiologic studies.

What is different about a molecular epidemiologic study and an epidemiologic study that uses molecular techniques is that there is a true merger of the disciplines: molecular biologic techniques integrated into epidemiologic studies, and epidemiologic methods fruitfully applied in a laboratory setting. When applied in epidemiologic studies, molecular biologic techniques enhance measures of diagnosis, prognosis, and exposure, reducing misclassification and increasing the power of studies to understand the etiology. When epidemiologic methods are applied in a laboratory setting, the focus on representative samples and population distributions illuminates the heterogeneity of microbial populations, and of human immune response to those populations, leading to more nuanced interpretations of results from 'model' organisms. A single molecular epidemiology course cannot even begin to substitute for the years of training required to become an expert in molecular techniques or

epidemiology; however, the student can develop sufficient vocabulary to begin bridging the two fields.

It is difficult for an individual who has never worked in a laboratory to appreciate the inherent variations that can occur in laboratory work. Instruments are calibrated and re-calibrated, and repaired or replaced; new lots made or purchased or reagents ordered from different companies; and specimens frozen and then thawed out for testing. Improper calibration, subtle changes in results with instrument repair or replacement or using new reagents, and variation in results with storage length or conditions, along with the human element, all contribute to variations in results. The extent that these are a problem depends upon what is being measured and how well it is known to associate with the underlying construct. Epidemiologic audiences often take molecular techniques at face value; the instructor must help them appreciate the complexity of molecular techniques and potential for error so they will scrutinize laboratory results with the same jaundiced eye applied to standard epidemiologic measures collected via questionnaire, from medical records, or via physical examination.

Similarly, an individual who has never collected data or managed large datasets has little appreciation for the complexities of data collection, management, and analysis. Collecting data and analysing results from a well-designed experiment in the laboratory is generally straightforward as compared to the field; in the laboratory, many of the sources of variation inherent in observational studies can be controlled by the investigator. By contrast, collecting data from people who live freely as opposed to caged animals poses many challenges: a participant may refuse to participate in part or all of a protocol; collecting data at multiple-study sites increases the chances that specimens are improperly labelled, shipped, or stored; and humans vary greatly in their medical history, diet, environmental exposures, and genotype in ways that are difficult to control. It is easy to ensure that a mouse has fasted for a specified time period before a specimen is collected; ensuring that a human participant has fasted is much more difficult.

Regardless of the student's previous training, the challenge to the student and instructor is to begin to truly bridge the different disciplines. Integrating molecular tools into an epidemiologic study not only increases the complexity of study conduct and analysis but can directly impact the study design. Measures may be time, space, or storage dependent: one test may be usefully applied anytime during a disease process, and another only at one stage in the natural history. Some tests work well on samples frozen for years, and others require almost immediate testing. This limits the options for working with databanks, the choice of study population, and the acceptability of testing to

participating study sites and participants. Further, determining these parameters often requires a study all by itself. Similarly, making population estimates of molecular parameters requires appropriately designed studies, something that is not merely a matter of common sense. Laboratorians often are not trained in population approaches and may not consider that any limitations of the study collections available to them from the freezer or their local clinician must be taken into account in interpreting the study results. It is still easy to find examples of studies limited in either of these manners in the literature for class discussion.

Finally, the availability of low-cost, high-throughput techniques, the omics, makes it possible to include these measures in epidemiologic studies. The results of these assays are (usually) multidimensional data that require considerable data management and analysis—often with specialized software—to reduce to measures amenable for integration with epidemiologic data. Genome-wide analyses of microbial and human genomes, assessments of changes in gene expression in health and disease or varying by exposures, and metabolomics screens are just a few examples.

The application of molecular tools in epidemiology is not new, particularly for infectious diseases. However, the number of molecular tools available and the range of potential applications have grown exponentially over the past two decades. The challenge to the instructor is to impart general principles that will hold, regardless of the techniques used. A molecular epidemiology course should familiarize the student with the underlying concepts and jargon of both molecular biology and epidemiology and build up appreciation for the strengths and develop understanding of the weaknesses of both molecular biology and epidemiology. In this chapter I outline my suggestions for core topics to cover in a molecular epidemiology course.

Teaching objectives

Courses in molecular epidemiology are offered not only in epidemiology departments but also in departments of the biological sciences. These are quite different audiences but central to teaching in either environment is to impart the strengths and weaknesses of both molecular and epidemiologic techniques. In a molecular epidemiology course preconceived notions regarding the validity and reliability of laboratory data and simplicity of epidemiologic methods often must be overcome before the students appreciate how choice of a particular molecular tool can influence the study design, conduct, and analysis and thus the study findings. Ideally, the student will have an opportunity at least to observe a laboratory setting and will be presented with exercises that lead them

to wrestle with the complexities of study design while incorporating molecular measuring tools. My teaching objective for an introductory course is to have students acquire a basic understanding of how to choose, evaluate, and apply new technologies in epidemiologic studies and to be able to evaluate the ethical implications of those choices.

In the absence of an entire course, students might be usefully introduced to the types of molecular tools that are increasingly applied in epidemiologic studies. Increasingly, a lecture or two on molecular epidemiology is included in courses on infectious disease epidemiology and cancer epidemiology. The former generally focuses on molecular typing tools used in outbreak investigation and surveillance, and the latter on biomarkers, describing the application of and enhancements gleaned from using more sensitive and specific measuring tools.

Regardless of whether taught as a unit in an existing course or as a separate course, including numerous examples of the benefits and limitations of making an epidemiologic study 'molecular' are in order. Two of the many past triumphs of molecular epidemiology were increased understanding of hepatitis etiology, leading to the prevention of post-blood-transfusion hepatitis, and the role of biomarkers in discovering the etiologic pathway linking benzene with haematologic cancers (Vineas and Perrera 2007). Molecular tools enhanced classification, enabled the identification of intermediate stages of pathogenesis, were essential in demonstrating a cause–effect relationship, and, for hepatitis, in developing both the hepatitis A and B vaccines and a test to screen blood for the hepatitis viruses, protecting the blood supply (Tobler and Busch 1997). Currently, molecular tools are a standard component of surveillance reports for foodborne illness and other infectious diseases. Most reports of outbreak investigations include molecular evidence confirming (or refuting) the epidemiologic evidence pointing to a disease reservoir or a transmission chain. Studies using expression arrays to compare gene expression in health and disease, or metabolomics screens to compare metabolic profiles in health and disease, are increasingly common for infectious and non-infectious conditions: a PubMed search for 'gene expression profile and disease' resulted in 6,409 citations and a search for 'metabolomics profile and disease' produced 151 citations (searches conducted 20 April 2013).

Teaching content and format

A model course syllabus for an introductory course is shown in Table 13.1. The presentation in this section follows the syllabus outline; suggested readings are also listed in Table 13.1.

Table 13.1 Course syllabus: suggested topics and associated readings

Date	Topic	Readings
1	Introduction: demonstrate range of potential applications of molecular tools to epidemiologic studies	Chapters 1 and 2, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.
		Chapter 1, Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.
2	Highlight how molecular tools increase understanding of the epidemiology infectious and non-infectious diseases	Chapter 3, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.
		Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.
		Dave, M., Higgins, P. D., Middha, S., and Rioux, K. P. (2012) The human gut microbiome: current knowledge, challenges, and future directions. *Translational Research*, **160**: 246–57.
3	Overview of molecular techniques	Chapter 5, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.
		National Library of Medicine (US). (2014) *Genetics Home Reference*. <http://ghr.nlm.nih.gov/handbook>, accessed 5 December 2014.
4	Techniques and their applications	Chapters 6 and 7, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.
		Chapter 6, Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.
5	Conceptual models and hypotheses	Lipsitch, M. Singer, R. S., and Levin, B. R. (2002) Antibiotics in agriculture: when is it time to close the barn door? *Proceedings of the National Academy of Sciences USA*, **99**: 5752–4; Smith, D. L., Harris, A. D., Johnson, J. A., Silbergeld, E. K., and Morris, J. G. Jr. (2002) Animal antibiotic use has an early but important impact on the emergence of antibiotic resistance in human commensal bacteria. *Proceedings of the National Academy of Sciences USA*, **99**: 6434–9.
6	Study design	Chapter 9, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.

Table 13.1 (continued) Course syllabus: suggested topics and associated readings

Date	Topic	Readings
7	Study design	Chapter 14, Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.
		Rundle, A. G., Vineis, P., and Ahsan, H. (2005) Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiology Biomarkers Preview*, **14**: 1899–907.
8	Measurement error	Chapter 8, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.
		Chapter 8, Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.
9	Validity and reliability	Schulte, P. A. and Perera, F. P. (1993) 'Validation', in P. A. Schulte and F. P. Perera, eds, *Molecular Epidemiology*. San Diego, CA: Academic Press, pp. 79–108.
10	Technical versus biologic variability	Vineis, P., Schulte, P. A., and Vogt, R. F. (1993) 'Technical variability in laboratory data', in P. A. Schulte and F. P. Perera, eds, *Molecular Epidemiology*. San Diego, CA: Academic Press, pp. 109–35.
11	Sample size calculation	Lai, D., King, T. M., Moyé, L. A., and Wei, Q. (2003) Sample size for biomarker studies: more subjects or more measurements per subject? *Annals of Epidemiology*, **13**: 204–8.
12	Quality control and assurance	Meinert, C. L. (1986) *Clinical Trials: Design, Conduct, and Analysis*, *Monographs in Epidemiology and Biostatistics*, vol. 8. New York: Oxford University Press, pp. 166–76.
		Chapter 10, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.
		Chapter 3, Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.
13	Human subjects concerns, biosafety	Chapter 12, Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology.* Amsterdam: Academic Press.

(*continued*)

Table 13.1 (continued) Course syllabus: suggested topics and associated readings

Date	Topic	Readings
14	Ethical concerns	Hawkins, A. K. and O'Doherty, K. C. (2011) "Who owns your poop?": insights regarding the intersection of human microbiome research and the ELSI aspects of biobanking and related studies. *BMC Medical Genomics*, **4**: 72.
		Chapter 2, Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer.

Knowledge of molecular methods that can be applied in epidemiologic studies, and their strengths and weaknesses

Reports of new molecular tools or technological improvements to existing tools appear almost daily in the literature. Fortunately, if one is familiar with the basic classes of tools, this myriad becomes understandable, enabling a more informed evaluation of the applicability and suitability of applying a particular technique in an epidemiologic context. Tools can be classified by underlying technique: the polymerase chain reaction (PCR), gel-based techniques, sequencing, antibody–antigen reactions, and hybridization. They can also be classified by the products measured: DNA, RNA, protein, and metabolites. Depending on how and what is measured, the measurement can be either a direct or indirect measure of the item of interest. Whether a tool is suitable for application in an epidemiologic context depends on the measures' reliability and validity, the type of specimens that must be tested, how those specimens must be collected, processed, and stored, and the costs. There are several good texts and web-based resources reviewing the different molecular techniques; the challenge is to find a resource that presents the material at the appropriate level for the intended audience. Unfortunately, it is more difficult to find information on the reliability and validity, and specimen collection and handling requirements, which are critical elements for choosing between techniques; at this writing, it is often necessary to use a variety of sources, from the literature to package inserts, to expert opinion to obtain this information.

The material must be presented at the correct level for the audience. While most students in biology will be familiar with at least some of the techniques, many epidemiology students will have rudimentary knowledge at best of the techniques or even of the biological products that might be measured. As the technical aspects remain a moving target, technical material is often best covered by a local expert.

The instructor can then present examples of applications in an epidemiologic context, and compare-and-contrast techniques by raising questions regarding the reliability and validity, and specimen collection and handling requirements. It is helpful to use a mix of historically notable and current examples.

How much time should be spent on describing the techniques depends on the background of the audience. My teaching objective is to enable the student to be able to understand the theory behind the technique at a level sufficient so that s/he can ask the right questions to determine if the technique will be appropriate for the desired purpose.

Choosing the right measuring tool for the application

What is the correct measuring tool depends on the research question, the current understanding of the etiology or underlying biology relating to the research question, and the study design. Depending on the research question, the study design may define the measuring tool, or the measuring tool the study design. Thus, it is best to include discussions about selecting the measuring tool(s) early on in the study design process.

With the advent of new measuring techniques, the development of the technique is often a rationale for conducting the study. However, just because a new technique is available does not mean it is the correct measuring instrument to address the research question. It is also useful to ask whether it is either necessary or appropriate to include a molecular measurement. For example, consider that exposure to cigarette smoke can be assessed by measuring smoking metabolites in the blood or urine; however, this assay gives no information about past exposure—which can be accurately assessed via questionnaire. If smoking history rather than current smoking exposure is the variable of interest, the investigator should use a questionnaire.

To emphasize these points, I include an in-class group exercise that requires students to compare and contrast using two different measuring tools for measuring approximately the same construct, in terms of strengths and weaknesses, validity and reliability, and requirements for specimen handling, processing, and testing, including costs. This compare-and-contrast exercise is also a component of the final class project.

Designing a validation and reliability study

Every molecular tool must be evaluated in the hands of the investigative team prior to implementation in the field, regardless of whether there are commercial kits available or published reports on the validity and reliability. Like cake mixes, laboratory kits are more or less sensitive to minor variations in instruments used, the 'standard' reagents added (such as impurities in the water),

accuracy of measurement, and processing. The investigator must be cognizant of this fact and feel comfortable overseeing appropriate testing of molecular techniques. For example, before going into the field to collect data, the investigator must consider how specimens are collected, what media they are collected into, the temperature they should be held at before and during shipping, whether specimens must be processed immediately or can be stored and batch processed at a later time, and so forth. If there are two or more technicians, both inter- and intra-technician variability should be assessed. Thus, it is essential that the student be comfortable designing and analysing validity and reliability studies and can identify the different situations where this type of study should be conducted.

In this section of the course, I generally review the differences between validity and reliability and the measures used to assess them. We discuss the components of the study design, particularly choosing samples for study that cover the range of values, and estimating the sample size required. We review current examples in the literature, and students work in small groups to design a specific example. The final class project includes descriptions of validity and reliability assessments for the selected molecular measure.

Designing a quality control/quality assurance plan

Molecular epidemiologic studies increase the complexity of data collection and processing. All epidemiologic studies should have in place methods to ensure that data meet a specific standard of quality (quality control), and ongoing procedures to monitor, verify, and document performance (quality assurance). Quality control and quality assurance procedures often appear in epidemiologic courses on clinical trials but coverage may be minimal otherwise. For clinical trials, the emphasis is often on questionnaire and specimen collection rather than on laboratory testing. By contrast, someone trained as a medical technologist will have substantial training on ensuring quality control and quality assurance with respect to laboratory testing but the training probably does not extend to issues beyond the laboratory.

In molecular epidemiologic studies, quality control and quality assurance procedures must encompass all aspects of data collection and analysis: from overseeing interviewers and documents, to specimen collection, handling, processing, and testing, to data entry, to analysis. Specimens must be tracked through various stages of processing: one specimen may become several and all resulting specimens must be correctly labelled, and their location known. If specimens are stored for testing in batches, the investigator must monitor variation within and between runs, determine the effects of storage on results, and institute procedures to minimize errors.

This material can often seem trivial. In order to bring the complexity to the fore, I ask students to list all the steps from data collection through analysis for a study of their own design and to note any step where either quality assurance or quality control is indicated. I then ask them to outline all the steps in specimen handling, list the sources of potential errors, and design appropriate quality control and quality assurance procedures for at least that one aspect of the proposed study. This exercise illuminates both the need to design quality control and quality assurance procedures up front and the time and effort involved (an example of this is included in Table 5, chapter 10, in Foxman 2012).

Familiarity with data repositories, databases, and associated software output specific to using molecular data

Molecular techniques, particularly PCR, make it possible to detect exquisitely small amounts of genetic material. While the technique has definite limitations, it does provide new options for screening existing data repositories or stored material collected for other reasons. For example, with PCR, it is possible to determine the microbial genetic content of stored Gram stains from vaginal smears and to determine human genetic content from small amounts of stored blood and blood products, saliva, urine, and other human specimens. Repositories are thus an important resource for conducting studies or testing new hypotheses in an economical fashion. They are also useful for obtaining control specimens for studies or conducting reliability or validity experiments. The value, however, depends on epidemiologic concerns: how the specimens were collected. Thus, an introduction to existing biologic specimen repositories and collections available from governmental agencies and foundations should also raise epidemiologic concerns. Students should be encouraged to consider if they might be able to address a specific research question using existing data, and the strengths and weaknesses of that approach, given a specific molecular measure.

There are a number of databases specific to molecular data. Using these data, whole genomes or genetic sequences can be aligned and compared, and guesses made as to what type of protein is coded for and its putative function. Using sequence from highly conserved genes (usually genes coding for the ribosome, which is present in all cells), the microbes present in a sample from a microbial community can be determined (e.g. determine what microbes are present in saliva). Databases can be searched for homologous genes appearing among other eukaryotic or prokaryotic genomes to gain insight into function. Currently, finding these resources can be difficult, even with excellent search engines, as appropriate key words are essential; thus, the goal is to make students aware that such databases might exist and point them to strategies for

finding them. A compendium of useful databases for molecular biology can be found in the Molecular Biology Database Collection, which is updated annually (for the 2013 listing, see Fernández-Suárez and Galperin 2013). Some of these databases are searchable using PubMed.

There has been tremendous advance in how data from these data-mining exercises are displayed. Depending on the class, the instructor may wish to include instruction as to how to align or search genetic material or how to interpret the output from common searches, such as BLAST, or how to create and read dendrograms.

Biosafety and protection of human subjects

Rules and regulations abound that relate to the collection, storage, handling, and shipping of biological specimens. There are additional rules and regulations relating to the use of recombinant DNA and infectious agents, including requirements to register with an institutional biosafety committee and to obtain the proper paperwork for shipping and receiving infectious materials. Investigators are also responsible for the health and safety of study personnel and of study participants. Ignorance of regulations is no defence in the eyes of the law; thus, we have failed our students if we do not introduce them to the rationale behind the regulations, and the various methods of oversight currently in place.

My goal for this section is to make students aware of the various situations where they must interact with regulatory bodies. For example, in the US, laboratories are inspected to ensure the health and safety of employees; inspection includes review of a laboratory manual that must be created according to specification. If a study uses biological specimens, the investigator must register with the biosafety committee, and ensure that the specimens are being handled in a laboratory that has the requisite level of biosafety. Personnel handling potentially infectious agents must be trained in proper precautions and offered vaccination. Some infectious agents are subject to specific regulation because of their potential use as bioweapons. Proper consent must be obtained from study participants for all anticipated uses of their specimens. The list is long, and discussing regulations is fatiguing. Requiring students to investigate and prepare a report on what regulations must be followed for conducting a particular study makes the effort both more real and less tedious.

Ethical issues

Melding fields also means melding cultures. This can lead to conflict, particularly with respect to areas that vary by culture. Ethical issues that should be addressed are authorship, protection of human subjects, sharing of data, and peer review.

All researchers conducting human and animal studies should know the regulations for protecting human and animal subjects, and the justification for those regulations. Two issues specific to molecular epidemiologic studies are (1) the use of specimens from biorepositories and (2) protecting confidentiality, when—if linked to appropriate databases—the results are uniquely identifying (such as genetic sequence). Molecular epidemiologic studies often take advantage of biorepositories or collect specimens that will constitute a future biorepository. While somewhat a moving target, there is increasing consensus on the need to acquire specific consent for inclusion in a biorepository at time of collection. Further, the use of already collected samples for new studies may require additional review by the appropriate regulating body, and possibly consent by the participants. Any sample that contains human genetic material or other material that uniquely identifies an individual (e.g. human microbiome studies) is held to a higher level of ethical protection. Students should be made sensitive to ethical concerns regarding collecting for and using samples from biorepositories and be made aware of the rapid changes in regulations and ethical guidelines relating to use of these data for genetic assays.

In most laboratory sciences the most prestigious position on an authorship list is the last one; this may or may not be true for a specific epidemiologic specialty. Students can be taught how to decipher who is the most senior author (e.g. who the corresponding author is or whose grant funded the project) on papers. They should also be made aware of requirements for authorship, at least as stipulated by many major journals. As interdisciplinary studies merge fields that may have different expectations for the amount of work that constitutes authorship, some discussion of case examples, and how authorship conflicts might be avoided, is helpful. The discussion should usefully distinguish between authorship and an acknowledgement; students should be cautioned that any person they intend to acknowledge should be consulted first.

The Human Genome Project set a standard for data sharing that includes uniform standards of reporting, and depositing of gene sequences in public databases on a timely basis. This level of transparency has been, and remains, quite rare in other epidemiologic contexts, although there is an increasing push towards the development of public access databases for large projects. The microbial community also has a history of sharing materials, as microbes are often renewable resources. The expectation upon publishing in some microbial journals is that all microbial isolates will be made available upon request. What types of materials should be made available, and over what timing, is a fruitful area for in-class discussion.

Finally, students should be introduced to the process of peer review, and a discussion of giving and receiving criticism. While peer review is not unique to

molecular epidemiology, peer review that crosses disciplines can be particularly, albeit unintentionally, harsh. Different cultures have different expectations regarding what points are important and which ones are assumed (or trivial). In a study of antibiotic resistance, the prevalence of the resistance phenotype has more clinical relevance than the prevalence of the specific gene(s) coding for resistance. However, reporting phenotype alone is considered trivial in some contexts, resulting in a less than favourable review. The ability to give constructive criticism and to interpret criticism constructively is a learnt skill. Thus, throughout the course, I have students criticize each other's work as a formal graded assignment.

Evaluation

An introductory course in an interdisciplinary area requires a mixture of different types of learning. A certain fraction will be knowledge based: learning the vocabulary of molecular biology and epidemiology. This is most easily evaluated by a test. Other aspects require analysis and synthesis, which is best evaluated by writing papers. I recommend evaluation at several points. I ask students to complete a quiz on the first day of class—or prior to class on a website—in order to assess the familiarity of the students with molecular and epidemiologic jargon. Students with a weaker understanding can be directed to materials to make up their deficits or to take a prerequisite. I also test knowledge using an in-class examination after covering the molecular tools and reviewing epidemiologic study designs. This ensures that all students have attained a similar level of comfort with the vocabulary of both fields, enhancing in-class discussions and work on the group project.

To help students bridge the fields, I assign group projects, which entail a literature review and then application of each of the topics covered in class to the selected topic. I have tried a mix of student-generated topics and asking them to choose from specific research questions. The former works better for more senior students. Regardless of how the project topic is selected, having a defined research question and condition that they learn well provides a structure for applying and wrestling with complex methodological considerations arising from melding molecular biology with epidemiology. I prefer group projects because discussion enhances understanding. However, to evaluate individual contribution, some of the preliminary papers towards the final project are developed individually. I review the final written project in draft and final form, and the proposal is also peer reviewed. In addition, each group presents their project orally for class discussion; the oral presentation is reviewed by me and by their peers.

Conclusion

Although the term 'molecular epidemiology' was coined in the early 1970s (see for a review Foxman and Riley 2001), the parameters of the field have only recently been defined. Laboratory methods are integrated into various aspects of infectious disease, cancer, cardiovascular, and environmental epidemiology; therefore, a course focusing on the vocabulary and principles needed to conduct these interdisciplinary efforts is a welcome addition to the curriculum of most epidemiologic programmes. By choosing appropriate examples, I hope that the curriculum described above can be adapted to emphasize the application in any epidemiologic subspecialty using molecular measures.

Bibliography

Fernández-Suárez, X. M. and Galperin, M. Y. (2013) The 2013 *Nucleic Acids Research* Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Research*, **41**: D1–7.

Foxman, B. and Riley, L. (2001) Molecular epidemiology: focus on infection. *American Journal of Epidemiology*, **153**: 1135–41.

Tobler, L. H. and Busch, M. P. (1997) History of post transfusion hepatitis. *Clinical Chemistry*, **43**: 1487–93.

Vineis, P. and Perera, F. (2007) Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiology Biomarkers and Prevention*, **16**: 1954–65. Erratum in (2007) *Cancer Epidemiology Biomarkers Prevention*, **16**: 2797.

Suggested texts

Carrington, M. and Roelzel, A. R., eds. (2001) *Molecular Epidemiology*. New York: Oxford University Press. This introductory text gives an overview of a variety of typing methods used for infectious diseases and some techniques used on the human genome and discusses some analytic techniques, e.g. phylogenic analysis. Presumes knowledge of biology; includes protocols.

Caugant, D.A., ed. (2009) *Molecular Epidemiology of Microorganisms: Methods and Protocols*. New York: Humana Press. Provides of an overview of common molecular genotyping methods, with laboratory protocols and examples of their application to investigations of infectious disease.

Dale, J. W. and von Schantz, M. (2002) *From Genes to Genomes: Concepts and Applications of DNA Technology*. New York: John Wiley & Sons. This text provides an introduction to the concepts and applications of molecular biology. A good reference for understanding a variety of basic molecular genetic techniques, including PCR, DNA sequencing, and gene expression.

Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology*. Amsterdam: Academic Press. This introductory text presents the key points of consideration when integrating molecular biology and epidemiology, discusses how using molecular tools in epidemiologic research affects program design and conduct, considers the ethical concerns that arise in molecular epidemiologic studies and provides a context for

understanding and interpreting scientific literature as a foundation for subsequent practical experience in the laboratory and in the field.

Morand, S.,Beaudeau, F., and Cabaret, J., eds. (2012) *New Frontiers of Molecular Epidemiology of Infectious Diseases*. London: Springer. This text focuses on the incorporation of population and evolutionary genetics into molecular epidemiology studies of infectious diseases.

Rothman, N., Hainaut, P., Schulte, P., Smith, M., Boffetta, P., and Perera, F., eds. (2011) *Molecular Epidemiology: Principles and Practice*. Lyon: International Agency for Research on Cancer. This text focuses on use of molecular tools in environmental epidemiology. The text has units on practical aspects of biomarkers, assessment of environmental exposures, integration of biomarkers into epidemiologic study designs, and applications of biomarkers to disease.

Schulte, P. A and Perera, F. P., eds. (1993) *Molecular Epidemiology: Principles and Practices*. San Diego: Academic Press, Inc. The first half of this text is on epidemiologic methods addressing specific concerns of molecular epidemiologic studies. Much of this is still relevant, although the examples are old. The second half consists of a series of chapters giving applications from a variety of different substantive areas, focusing primarily on chronic illnesses and the use of biomarkers. The examples are updated in the second edition.

Chapter 14

Social inequalities in health

Nancy Krieger

Introduction to social inequalities in health

Teaching about social inequalities in health is fundamental to epidemiology. The first reason is substantive: social injustice harms health (Krieger 1999; WHO Commission on the Social Determinants of Health (CSDH) 2008; Krieger 2011). Understanding how we embody inequality, and what can be done to prevent this, is key to the mission of both epidemiology and public health overall. The second reason is methodological: health inequities affect the conduct of rigorous science, whether or not the focus of the research is on social inequalities in health per se (Krieger 2011). Epidemiologic evidence, whether experimental or observational, can be rendered invalid by socially patterned selection bias, confounding, and misclassification (Davey Smith 2003; Krieger 2007a). We ignore these issues at our peril—with life-and-death consequences for the public's health.

In this chapter, I offer guidelines for teaching an introductory epidemiological course on social inequalities in health. My approach is partly based on teaching, for the past twenty years, a US graduate-level public health course on 'History, politics, and public health: theories of disease distribution and social inequalities in health' (Krieger 2011). Also relevant is my etiological, methodological, and theoretical work as a social epidemiologist concerned with analysing, monitoring, and addressing health inequities. My perspective, which emphasizes critical learning, is grounded in ecosocial theory. This theory of disease distribution, which I first proposed in 1994 (Krieger 1994) and have elaborated since (Krieger 2001a, 2011), is fundamentally concerned with health inequities, the relevance of history, biology, and society for epidemiologic thinking, and socially responsible science.

Teaching objectives

The potential scope of an introductory course on social inequalities in health is enormous, given the range of substantive, conceptual, and methodological

topics that could be covered. Moreover, a class on this topic could be taught not only to graduate students in the health professions but also in other disciplines (e.g. sociology, anthropology, policy, etc.) and to students at other levels (e.g. undergraduate). It could, likewise, be designed for health activists and community members working to eliminate health inequities (see e.g. International People's Health University 2013 and NACCHO 2013). Because *Teaching Epidemiology is* intended for 'teachers in epidemiology, public health, and clinical medicine', however, my chapter proposes a graduate-level, sixteen-session introductory course.

As a starting point, when teaching epidemiologists and other health professionals, I have found it more fruitful pedagogically to work with students first for conceptual clarity on the substantive questions (e.g. what are health inequities and what causes them), and then address the methodological issues raised by trying to answer these questions empirically and testing the relevant hypotheses. For this reason, I suggest that an introductory course on social inequalities in health focus primarily on definitional and conceptual issues and secondarily on methodological concerns. A follow-up course on methods for studying health inequities would be a logical successor.

Based on these considerations, Box 14.1 provides a sample course description and set of learning objectives. Because the proposed class is meant to be global in reach and easily adaptable to whatever country-context in which it is taught, the course description does not focus on a particular country or geographic area. Where appropriate, however, case examples should be drawn from the country focus of the course, to bring home the issue of health inequities to the enrolled students.

It would, of course, be possible to design separate introductory epidemiologic courses for any of the areas addressed by the proposed class, for example, on theories of disease distribution for studying health inequities, on class and health, or racism and health, or gender and health, or sexuality and health, or the geography or history of health inequities, etc. While these more focused types of introductory courses are needed (and do exist), an introductory course that integrates material on a wide range of health inequities, to make clear their interconnections, similarities, and differences, is essential. After all, our bodies daily integrate and embody our societal and ecological context (Krieger 1994, 2011); our teaching (and research) should do no less.

The learning objectives provided in Box 14.1 build on the course description and clarify the specific knowledge and skills that students should obtain as a result of having taken the class. The emphasis on students gaining a critical perspective and acquiring the capacity to engage in debates stems from my pedagogic orientation. Students learn best when they are encouraged to be

Box 14.1 Course description and learning objectives for proposed introductory course on social inequalities in health

Course description:

This sixteen-session course is an introduction to the epidemiology of social inequalities in health. It is intended to provide a critical understanding of what health inequities are and why they matter—for the lives of those burdened by health inequities and for the rigour of epidemiologic research. The course will

1. introduce definitions of and debates over the meaning of 'social inequalities in health';
2. briefly review historical dimensions of health inequities, to give context to current trends;
3. introduce key theoretical perspectives that guide epidemiologic research on health inequities (e.g. social production of disease/political economy of health, neo-materialist, psychosocial, life course, and ecosocial);
4. define key dimensions of health inequities within and between countries, especially in relation to class, racism, gender, sexuality, and global politics;
5. critically review epidemiologic research on these different—and interconnected—dimensions of health inequities, taking into account etiologic pathways by which inequality is embodied, in relation to both level and life course; and
6. consider debates over whether it is 'politically correct'—or correct and necessary—for epidemiologists to pay attention to links between social inequality and health, as either a primary research focus or in investigations not directly concerned with health inequities.

Learning objectives:

The overall goal is for students to develop a critical understanding of 'social inequalities in health' and why they matter. By the end of the course, students will be able to do the following:

1. Define what is meant by 'social inequalities in health', and discuss debates about its meaning.
2. Describe different theoretical frameworks epidemiologists use to study health inequities.

Box 14.1 Course description and learning objectives for proposed introductory course on social inequalities in health *(continued)*

3. Describe key aspects of different domains of health inequities covered, singly and combined: who is affected, compared to whom? What are the trends over time, overall and for specific outcomes? What are the pathways of embodiment? And how can epidemiologists measure and analyse the relevant exposures, in relation to both level and life course?
4. Debate whether epidemiologists need to be concerned about health inequities on both substantive and methodological grounds—if so, why; if not, why not.

active and questioning learners, not passive consumers of received knowledge (Friere 1970). The real world of scientific inquiry is fraught with unanswered questions, conceptual and methodological debates, and contending perspectives. Students need to be trained to enter this world, equipped with relevant conceptual tools, substantive knowledge, and the ability to challenge dogma and debate ideas.

Teaching content

The course content, presented in outline form in Box 14.2 and described in more detail below, flows from the course objectives. Table 14.1 provides a brief guide to key concepts the course should cover, and Table 14.2 lists fifteen suggested readings.

Session 1: what are 'social inequalities in health'? Definitions and debates

This first session should address what is even meant by the phrase 'social inequalities in health' (see Box 14.1). It should likewise clarify that because the focus of the course is on health inequities—that is, how social inequality shapes population health—other causes of disease and disease distribution unrelated to inequity will be discussed only as warranted. A key concept is that the non-equivalence of health status between groups can arise in two very different ways: (1) because they are socially produced and are due to unfair and unjust societal conditions, or (2) because they arise from variations that are not socially determined (see Box 14.1; Whitehead 1992; Krieger 2005a; WHO CSDH 2008). The former constitute what is increasingly referred to as 'health inequities' (also termed 'social inequalities in health'); the latter, simply differences in health

Box 14.2 Outline for proposed introductory epidemiologic course on 'social inequalities in health'

Session 1	What are 'social inequalities in health'?—Definitions and debates
Session 2	Health inequities in historical perspective: a brief review
Session 3	Theoretical frameworks for epidemiologic research on health inequities
Session 4	Key dimensions of health inequities within and between countries: global politics, class, racism, gender, and sexuality, in context
Session 5	History, levels, life course, and pathways of embodiment leading to health inequities—and the fallacies of 'nature versus nature'
Session 6	Health inequities and social class: pathways and measurement
Session 7	Health inequities and racism: pathways and measurement
Session 8	Health inequities and gender: pathways and measurement
Session 9	Health inequities and sexuality: pathways and measurement
Session 10	Health inequities between countries and regions: pathways and measurement
Session 11	Case example: epidemiologic analyses of health inequities for a particular population
Session 12	Case example: epidemiologic analyses of health inequities for a particular outcome
Session 13	'Politically correct' or correct science?—The case of racism versus 'race' and health inequities
Session 14	'Politically correct' or correct science?—The case of hormone therapy, cardiovascular disease, and breast cancer
Session 15	Implications of epidemiologic research on health inequities for the public's health
Session 16	Who and what is accountable for social inequalities in health: summation, student projects, and course wrap-up

status. For example, the fact that men get prostate cancer but women get cervical cancer is a difference, not an inequity. By contrast, which women and which men are at high risk of being diagnosed with or dying from these diseases *can* be a matter of inequity, as evidenced by racial/ethnic and class patterns for these health outcomes (Krieger 2005a). The readings assigned for this class should

Table 14.1 Short list of key concepts: epidemiology and health inequities.

Key concept (alphabetical order)	Brief explications, excerpted from 'A glossary for social epidemiology' (Krieger 2001b)
Biologic expressions of social inequality	'**Biologic expressions of social inequality** refers to how people literally embody and biologically express experiences of economic and social inequality, from *in utero* to death, thereby producing social inequalities in health across a wide spectrum of outcomes. . . .' [See also Krieger 2011]
Discrimination	'**Discrimination** refers to "the process by which a member, or members, of a socially defined group is, or are, treated differently (especially unfairly) because of his/her/their membership of that group" (Jary and Jary 1995: 169). This unfair treatment arises from "socially derived beliefs each [group] holds about the other" and "patterns of dominance and oppression, viewed as expressions of a struggle for power and privilege" (Marshall 1994: 125–6). . . . Predominant types of adverse discrimination are based on race/ethnicity, gender, sexuality, disability, age, nationality, and religion, and, although not always recognized as such, social class . . . Social epidemiologic analyses of health consequences of discrimination require conceptualizing and operationalizing diverse expressions of exposure, susceptibility, and resistance to discrimination . . . recognizing that individuals and social groups may be subjected simultaneously to multiple—and interacting—types of discrimination. . . .' [See also Krieger 1999, 2011; Williams and Mohammed 2009]
Ecosocial theory of disease distribution	'**Ecosocial [theory]** . . . seek[s] to integrate social and biologic reasoning and a dynamic, historical and ecological perspective to develop new insights into determinants of population distributions of disease and social inequalities in health. The central question for ecosocial theory is: "*who and what is responsible for population patterns of health, disease, and well-being, as manifested in present, past, and changing social inequalities in health*?" . . . Core concepts for ecosocial theory . . . include: (1) **embodiment** . . . (2) **pathways of embodiment** . . . (3) **cumulative interplay between exposure, susceptibility, and resistance** [across the life course] . . . (4) **accountability and agency** . . .' [See also Krieger 1994, 2001a, 2008, 2011]
Embodiment	'a concept referring to how we literally incorporate, biologically, the material and social world in which we live, from *in utero* to death; a corollary is that no aspect of our biology can be understood absent knowledge of history and individual and societal ways of living'. [See also Krieger 1994, 2005b, 2011]

Table 14.1 (continued) Short list of key concepts: epidemiology and health inequities.

Key concept (alphabetical order)	Brief explications, excerpted from 'A glossary for social epidemiology' (Krieger 2001b)
Gender, sexism, and sex	'**Gender** refers to a social construct regarding culture-bound conventions, roles, and behaviors for, as well as relations between and among, women and men and boys and girls. . . . **Sexism,** in turn, involves inequitable gender relations and refers to institutional and interpersonal practices whereby members of dominant gender groups (typically men) accrue privileges by subordinating other gender groups (typically women) and justify these practices via ideologies of innate superiority, difference, or deviance. Lastly, **sex** is a biological construct premised upon biological characteristics enabling sexual reproduction. . . . Sex-linked biological characteristics (e.g., presence or absence of ovaries, testes, vagina, penis; various hormone levels; pregnancy, etc.) can, in some cases, contribute to gender differentials in health but can also be construed as **gendered expressions of biology** and erroneously invoked to explain **biologic expressions of gender**. . . .' [See also Doyal 1995, Krieger 2003a, Payne 2006, Fausto-Sterling 2012]
Human rights and social justice	**Human rights,** as a concept, presumes that all people 'are born free and equal in dignity and rights' (United Nations 1948), and provides a universal frame of reference for deciding questions of equity and social justice. . . . Human rights norms are premised, in the first instance, upon the 1948 Universal Declaration of Human Rights (United Nations 1948) and its recognition of the indivisibility and interdependence of civil, political, economic, social, and cultural rights. A 'health and human rights' framework thus not only spurs recognition of how realization of human rights promotes health but also helps translate concerns about how violation of human rights potentially harms health into concrete and actionable grievances which governments and the international community are legally and politically required to address. Understanding of what prompts violation of human rights and sustains their respect, protection and fulfilment is, in turn, aided by **social justice** frameworks, which explicitly analyze who benefits from—and who is harmed by—economic exploitation, oppression, discrimination, inequality, and degradation of 'natural resources' . . . [See United Nations 1948; Boucher and Kelly 1998; Gruskin et al. 2005]
Life course perspective	'**Lifecourse perspective** refers to how health status at any given age, for a given birth cohort, reflects not only contemporary conditions but embodiment of prior living circumstances, *in utero* onwards . . .' [See also Kuh and Ben-Shlomo 2004; Davey Smith 2003]

(*continued*)

Table 14.1 (continued) Short list of key concepts: epidemiology and health inequities.

Key concept (alphabetical order)	Brief explications, excerpted from 'A glossary for social epidemiology' (Krieger 2001b)
Multilevel analysis	'**Multi-level analysis** refers to statistical methodologies, first developed in the social sciences, that analyze outcomes simultaneously in relation to determinants measured at different levels (e.g., individual, workplace, neighborhood, nation, or geographic region existing within or across geopolitical boundaries) . . .' [See also Diez-Roux 2002; Subramanian et al. 2003]
Poverty, deprivation (material and social), and social exclusion	'To be **impoverished** is to lack or be denied adequate resources to participate meaningfully in society. A complex construct, **poverty** is inherently a normative concept that can be defined—in both absolute and relative terms—in relation to: "need", "standard of living", "limited resources", "lack of basic security", "lack of entitlement", "multiple deprivation", "exclusion", "inequality", "class", "dependency", and "unacceptable hardship" (Gordon and Spicker 1999) . . . **Deprivation** can be conceptualized and measured, at both the individual and area level, in relation to: **material deprivation,** referring to "dietary, clothing, housing, home facilities, environment, location and work (paid and unpaid)", and **social deprivation,** referring to rights in relation to "employment, family activities, integration into the community, formal participation in social institutions, recreation and education" (Townsend 1993 : 93) . . . **Social exclusion,** another term encompassing aspects of poverty, in turn focuses attention on not only the impact but also the process of marginalization . . .' [See also Krieger et al. 1997; Shaw et al. 2007]
Psychosocial epidemiology	'A **psychosocial** framework directs attention to both behavioral and endogenous biological responses to human interactions . . .' [See also Marmot 2004]
Race/ethnicity and racism	'**Race/ethnicity** is a social, *not* biological, category, referring to social groups, often sharing cultural heritage and ancestry, that are forged by oppressive systems of race relations, justified by ideology, in which one group benefits from dominating other groups, and defines itself and others through this domination and the possession of selective and arbitrary physical characteristics (e.g., skin color). **Racism** refers to institutional and individual practices that create and reinforce oppressive systems of race relations (see "discrimination"). **Ethnicity,** a construct originally intended to discriminate between "innately" different groups allegedly belonging to the same overall "race", is now held by some to refer to groups allegedly distinguishable on the basis of "culture"; in practice, however, "ethnicity" cannot meaningfully be disentangled from "race" in societies with inequitable race relations, hence the construct "race/ethnicity" . . .' [See also Banton 1998; Krieger 1999]

Table 14.1 (continued) Short list of key concepts: epidemiology and health inequities.

Key concept (alphabetical order)	Brief explications, excerpted from 'A glossary for social epidemiology' (Krieger 2001b)
Sexualities and heterosexism	'**Sexuality** refers to culture-bound conventions, roles, and behaviors involving expressions of sexual desire, power, and diverse emotions, mediated by gender and other aspects of social position (e.g., class, race/ethnicity, etc.). Distinct components of sexuality include: sexual identity, sexual behavior, and sexual desire. Contemporary "Western" categories by which people self-identify or can be labeled include: heterosexual, homosexual, lesbian, gay, bisexual, "queer", transgendered, transsexual, and asexual. **Heterosexism,** the type of discrimination related to sexuality, constitutes one form of abrogation of sexual rights and refers to institutional and interpersonal practices whereby heterosexuals accrue privileges (e.g., legal right to marry and to have sexual partners of the "other" sex) and discriminate against people who have or desire same-sex sexual partners, and justify these practices via ideologies of innate superiority, difference, or deviance . . .' [See also Parker and Gagnon 1995; Meyer and Northridge 2007; Committee on LGBT Health Issues 2011]
Society, social, societal, and culture	'**Society,** originally meaning "companionship or fellowship", now stands as "our most general term for the body of institutions and relationships within which a relatively large group of people live and as our most abstract term for the condition in which such institutions and relationships are formed" (Williams 1983 : 291). **Social,** as an adjective, likewise has complex meanings: "as a descriptive term for *society* in its now predominant sense of the system of common life," and also as "an emphatic and distinguishing term, explicitly contrasted with *individual* and especially *individualist* theories of society" [italics in the original] (Williams 1983: 286). **Societal,** in turn, serves as a "more neutral reference to general **social** formations and institutions" (Williams 1983: 294). By this logic, **social epidemiology** and its social theories of disease distribution stand in contrast to **individualistic epidemiology,** which relies on individualistic theories of disease causation . . . **Culture,** originally a "noun of process" referring to "the tending of something, basically crops or animals" (Williams 1983 : 87), presently has three distinct meanings: "(i) the independent and abstract noun which describes a general process of intellectual, spiritual, and aesthetic development . . . ; (ii) the independent noun, whether used generally or specifically, which indicates a particular way of life, whether of a people, a period, a group, or humanity in general; and . . . (iii) the independent and abstract noun which describes the work and practices of intellectual and especially artistic activity" (Williams 1983 : 90). In social epidemiology, meaning (ii) predominates . . .'[1]

(*continued*)

Table 14.1 (continued) Short list of key concepts: epidemiology and health inequities.

Key concept (alphabetical order)	Brief explications, excerpted from 'A glossary for social epidemiology' (Krieger 2001b)
Social class and socioeconomic position	'**Social class** refers to social groups arising from interdependent economic relationships among people. These relationships are determined by a society's forms of property, ownership, and labor, and their connections through production, distribution, and consumption of goods, services, and information . . . Class, as such, is not an a priori property of individual human beings, but is a social relationship created by societies. As such, social class is logically and materially prior to its expression in distributions of occupations, income, wealth, education, and social status. One additional and central component of class relations involves an asymmetry of economic exploitation, whereby owners of resources (e.g., capital) gain economically from the labor or effort of non-owners who work for them . . . **Socioeconomic position,** in turn, is an aggregate concept that includes both resource-based and prestige-based measures, as linked to both childhood and adult social class position. . . . The term "socioeconomic status" should be eschewed because it arbitrarily (if not intentionally) privileges "status"—over material resources—as the key determinant of socioeconomic position. . . .' [See also Krieger et al. 1997; Wright 1997; Shaw et al. 2007]
Social determinants of health	'**Social determinants of health** refer to both specific features of and pathways by which societal conditions affect health and that potentially can be altered by informed action. . . .' [See also WHO CSDH 2008; Krieger 2011]
Social inequality or inequity in health and social equity in health	'**Social inequalities (or inequities) in health** refer to health disparities, within and between countries, that are judged to be unfair, unjust, avoidable, and unnecessary (meaning: are neither inevitable nor unremediable) and which systematically burden populations rendered vulnerable by underlying social structures and political, economic, and legal institutions. . . . **Social equity in health,** in turn, refers to an absence of unjust health disparities between social groups, within and between countries. . . .' [See also Whitehead 1992; WHO CSDH 2008]
Social production of disease/political economy of health	'**Social production of disease/political economy of health** refer to related (if not identical) theoretical frameworks that explicitly address economic and political determinants of health and distributions of disease within and across societies, including structural barriers to people living healthy lives . . .' [See also Doyal 1979; Navarro and Muntaner 2004; Krieger 2011]

Table 14.1 (continued) Short list of key concepts: epidemiology and health inequities.

Key concept (alphabetical order)	Brief explications, excerpted from 'A glossary for social epidemiology' (Krieger 2001b)
Social production of scientific knowledge	'**Social production of scientific knowledge** refers to ways in which social institutions and beliefs affect recruitment, training, practice, and funding of scientists, thereby shaping what questions we, as scientists, do and do not ask, the studies we do and do not conduct, and the ways in which we analyze and interpret data, consider their likely flaws, and disseminate results . . .' [See also Ziman 2000; Krieger 2011]
Stress	'**Stress,** a term widely used in the biological, physical, and social sciences, is a construct whose meaning in health research is variously defined in relationship to "stressful events, responses, and individual appraisals of situations" (Cohen et al. 1995 : 3). Common to these definitions is "an interest in the process in which *environmental demands tax or exceed the adaptive capacity of an organism, resulting in psychological or biological changes that may place persons at risk for disease*" [italics in original] (Cohen et al. 1995 : 3) . . .'
Theories of disease distribution	'**Theories of disease distribution** seek to explain current and changing population patterns of disease across time and space and, in the case of social epidemiology, across social groups (within and across countries, over time) . . .' [See also Krieger 2001a: Krieger 2011]

Emphasis in the original

[1] The text in 'Society, social, societal, and culture' is adapted from Krieger (2001b).

review the global terminology on health inequities and the specific terminology employed in the country where the course is taught (e.g. in the US, the most widely used terminology refers to 'health disparities'; Carter-Pokras and Bacquet 2002; Krieger 2005a).

Session 2: health inequities in historical perspective—a brief review

The second session should provide students with an historical perspective on health inequities. This will set the basis for what the class will cover and also counter the false impression, common to students new to this field, that epidemiologists have only recently become aware of social inequalities in

Table 14.2 List of fifteen suggested readings for students interested in epidemiology and health inequities.

Topic area	References (in alphabetical order)
Social epidemiology and health inequities	Berkman, L., Kawachi, I., and Glymour, M., eds. (2014)*Social Epidemiology* (2nd edn). Oxford: Oxford University Press.
	Davey Smith, G., ed. (2003) *Health Inequalities: Lifecourse Approaches.* Bristol: Policy Press.
	Krieger, N. (1994) Epidemiology and the web of causation: has anyone seen the spider? *Social Science and Medicine*, **39**: 887–903.
	Krieger, N. (2001a) Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology*, **30**: 668–77.
	Krieger, N. (2011) *Epidemiology and the People's Health: Theory and Context*. New York: Oxford University Press.
	Oakes, J. M. and Kaufman, J. S., eds. (2006) *Methods in Social Epidemiology.* San Francisco, CA: Jossey-Bass.
	Young, T. K. (1998) *Population Health: Concepts and Methods.* Oxford: Oxford University Press.
Health inequities in context	Birn, A-E., Pillay, Y., and Holtz, T.H. (2009)*Textbook of International Health: Global Health in a Dynamic World* (3rd edn). New York: Oxford University Press.
	Evans, T., Whitehead, M., Diderichsen, F., Bhuiya, A., Wirth, M., and Whitehead, M., eds. (2001) *Challenging Inequities in Health: From Ethics to Action.* Oxford: Oxford University Press.
	Gruskin, S., Grodin, M. A., Annas, G. J, and Marks, S. P., eds. (2005)*Perspectives on Health and Human Rights*. New York: Routledge.
	Hofrichter, R., ed. (2003) *Health and Social Justice: Politics, Ideology, and Inequity in the Distribution of Disease.* San Francisco: Jossey-Bass.
	Kunitz, S. (2006) *The Health of Populations: General Theories and Particular Realities.* Oxford: Oxford University Press.
	Navarro, V. and Muntaner, C., eds. (2004) *Political and Economic Determinants of Population Health and Well-Being: Controversies and Developments.* Amityville, NY: Baywood Publishing Company.
	Porter, D. (1999) *Health, Civilization and the State: A History of Public Health from Ancient to Modern Times.* London: Routledge.
	World Health Organization Commission on the Social Determinants of Health (WHO CSDH). (2008) *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health*. Geneva: WHO.

health and that social epidemiology is a novel discipline. There are three key points:

1. Scholars have long recognized that ways of living and working affect health, with such observations found in the earliest known medical documents. Whether these differential risks are seen as unfair, however, depends on prevailing and contending views about causes of social inequality (e.g. innate or imposed; see Porter 1999; Krieger 2000, 2011).
2. The emergence of epidemiology as a scientific discipline in the early nineteenth century was inextricably bound to concerns about destitution, as spurred by the global public health impact of that era's massive transformations in ways of living—and of dying. The Industrial Revolution and unleashing of laissez-faire capitalism sparked the creation of a fast-growing, impoverished, urban working class, massive increases in international trade, and an expanding military presence in colonized countries and outposts across the five continents. These developments set the basis for unprecedented European epidemics of cholera and yellow fever, along with declining life expectancy, especially among the urban poor. Through the urgent study of these problems, epidemiology, as a self-designated field of scientific study, was born (see Porter 1999; Krieger 2000, 2007, 2011).
3. Health inequities are historically contingent, meaning that their magnitude and specific forms of expression (for particular outcomes and also for overall measures, such as premature mortality and life expectancy) depend on particular societal conditions (Kunitz 2006). Consider, for example, changes in the association between socio-economic position and smoking: during the twentieth century, in both the US and several European countries, cigarette-smoking rates initially were higher in professional compared to working-class occupations—a trend that then reversed itself (Graham 1996; Brandt 2007).

Yet, while it is important to grasp that there is no one simple 'story' of health inequities, a general statement still holds: (1) material resources and knowledge are needed to live a healthy life, (2) social inequality results in the unfair distribution of these resources, and (3) groups subjected to social and economic deprivation typically suffer the worst health status while groups who benefit from the social and economic systems producing these inequities tend to fare best (Krieger 2011). To highlight the relevance of these points for the students in the class, the session should include historical examples of health inequities and investigators who have researched them for the country focus of the course.

Session 3: theoretical frameworks for epidemiologic research on health inequities

The third session should provide an introductory overview of the different theoretical frameworks epidemiologists currently use to analyse health inequities. This is because theory determines what is studied and what is ignored, using which methods, with which interpretations (see Table 14.1; Krieger 1994, 2001a, 2011). Contending twentieth-century epidemiologic theories range from individualistic biomedical and lifestyle approaches, which emphasize individual-level biology (especially genetics) and choice as key determinants of health, to more contextualized frameworks concerned with societal determinants of health inequities (see Tesh 1988; Krieger 1994, 2001a, 2011). Among the latter, a central premise is that health inequities arise from unjust relationships between groups, not intrinsic characteristics (see Tesh 1988; Krieger 1994, 2001a, 2011).

Within these more contextualized approaches, some frameworks focus almost exclusively on the social determinants of health and leave biology relatively opaque (e.g. social production of disease/political economy of health, social determinants of health, and health and human rights; see e.g. Doya1 1979; Hofrichter 2003; Navarro and Muntaner 2004; Gruskin et al. 2005). Others are more biologically or psychologically oriented but do not systematically consider political economy (e.g. life course, psychosocial; see e.g. Kuh and Ben-Shlomo 2004; Marmot 2004). Still others such as ecosocial theory (see Table 14.1 and Fig. 14.1;

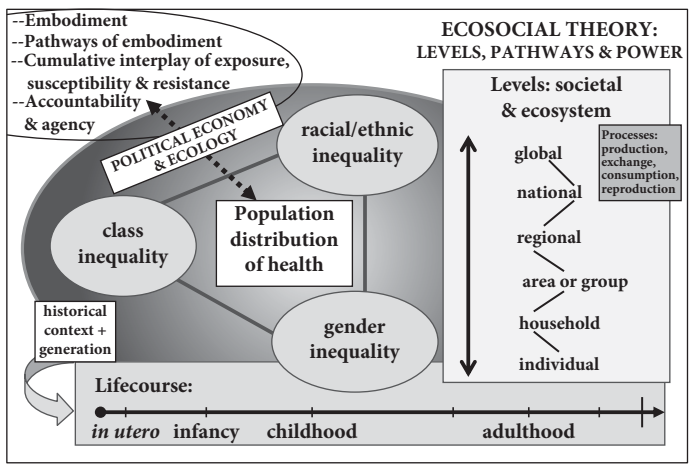


Fig. 14.1 Embodying inequality: an ecosocial approach to analysing disease distribution population health, and health inequities (Krieger 1994, 2001a, 2008, 2011).

Krieger 1994, 2001a, 2008, 2011) call for multilevel epidemiologic theorizing conceptualized in societal, biological, ecological, and historically contingent terms (see also Susser 1996). A central concern of ecosocial theory, for example, is how population distributions of and inequities in health, disease, and well-being constitute the embodied consequences of people's societal and ecologic context and hence both political economy and political ecology. Core constructs accordingly pertain to the process and pathways of embodiment and how they involve the interplay of social and biological exposures, susceptibility, and resistance across the life course, with issues of agency and accountability referring not only to who and what is responsible for disease distribution but also how epidemiologists and other scientists study and explain these distributions (Table 14.1 and Fig. 14.1). In the session discussion, students should analyse and debate the different types of etiologic hypotheses encouraged by each of these theoretical perspectives.

Session 4: key dimensions of health inequities within and between countries—global politics, class, racism, gender, and sexuality, in context

The fourth session would present a broad overview of the major types of social inequalities in health currently under investigation. These involve health inequities both within and between countries, principally in relation to global politics, class, racism, gender, and sexuality, in context (see Table 14.1). As revealed by a critical analysis of who and what makes populations (Krieger 2012a), at issue is how these observed population-level biological expressions of social inequality arise due to exploitative and/or oppressive relationships between the groups co-defined by their inequitable relationships, and the impact these inequities have on the material, psychosocial, and ecosystem conditions in which people live, ail, and die (Krieger 2011). Emphasis should be on a preliminary discussion of the interconnections, similarities, and distinctions between these various types of inequity and their implications for health, noting that each type receives more in-depth focus in sessions 6 to 10. Students interested in debates over the causes and manifestations of diverse forms of societal inequity should be referred to articles describing current and contending social science perspectives on these issues (e.g. as contained in such resources as the online *International Encyclopedia of the Social and Behavioral Sciences* (Smelser and Baltes 2004)).

Session 5: history, levels, life course, and pathways of embodiment leading to health inequities—and the fallacies of 'nature versus nurture'

This session would emphasize the historically contingent processes generating population health inequities, by considering levels, life course, and pathways of

embodiment, in historical context (see Table 14.1). Two distinct but related topics require consideration: (1) gene expression in societal and ecological context, and (2) temporal changes in the social patterning of adverse exposures. The first concerns the biological constructs of 'norms of reaction' (Lewontin 2000; Gilbert and Eppel 2009) and 'flexible phenotypes' (Piersma and van Gils 2011). At issue is how an organism's biophysical context (including active engagement with organisms in its own and other species) affects gene expression across the life course, such that any given genotype has the potential to produce a diversity of phenotypes (within the constraints imposed by evolutionary history; a bee will not suddenly develop into a tree); stated another way, 'bodies express ecology' (Piersma and van Gills 2011, p. 2). The contrast is to misleading and erroneously biological determinist pittings of 'nature versus nurture,' a framing which contemporary genetic and other biological research, especially in the field of evolutionary ecological developmental biology, show to be a deeply flawed approach to understanding biology (Lewontin 2000; Gilbert and Eppel 2009; Keller 2010; Fausto-Sterling 2012). For epidemiology, a consequently more productive framing is to conceptualize disease and other biological characteristics as 'emergent embodied phenotypes' (Krieger 2012a, 2013).

With regard to changing societal disease distributions, a useful example could be that of changing trends in the magnitude of health inequities in smoking-related diseases, taking into account exposures at the global, country-specific, local, community, and individual levels, and also whether exposure starts in utero, in childhood, or in adulthood (Graham 1996; Brandt 2007). Topics of discussion, to be considered in relation to both birth cohort and period effects, could include global trade agreements regarding the production, sale, and consumption of cigarettes; government policies about cigarette taxes and where smoking is allowed; tobacco industry efforts to target socially vulnerable smokers, and public health initiatives to counter these campaigns; and smokers' physiological addiction to, reliance on, and enjoyment of cigarettes for their social and psychoactive properties. Equally germane examples could include access to safe drinking-water (McMichael 2001; Whiteford and Whiteford 2005; United Nations Development Programme (UNDP) 2006), inequitable impacts of global climate change (McMichael 2001; McMichael and Butler 2011), or changing distributions of disease biomarkers (e.g. the breast cancer oestrogen receptor; Krieger 2013).

Session 6: health inequities and social class—pathways and measurement

The sixth session would focus on class inequities in health (see Table 14.1). Describing their magnitude and analysing their causes presumes an understanding of what social class is and how it can be empirically measured, at

different levels, across the life course (Krieger et al. 1993, 1997; Wright 1997; Shaw et al. 2007). Strengths and limitations of individual-, household-, and area-based measures of socio-economic position (e.g. income, education, occupation, wealth, debt) should be discussed, including how their meaning may vary by age, gender, race/ethnicity, and a country's economic level and the relative size of its formal, informal, and illegal economy (Krieger et al. 1997; Shaw et al. 2007). Assigned reading should include (1) review articles conceptualizing and measuring socio-economic position in epidemiologic research—and giving concrete examples of instruments employed—in relation to levels and life course, including the different pathways by which class inequality can be embodied; and (2) a selection of epidemiologic investigations, with at least some relevance to the country or region that is the focus of the class, and which do a good as well as poor job of analysing class inequities in health. The classroom discussion can then critique the specific articles, as informed by the more conceptual review articles; the same approach can be used for sessions 7 to 10. Discussion for sessions 6 to 9 should likewise consider how the measures of social position employed (e.g. class for session 6) can be used to stratify or 'control' for confounding by these social variables in studies not directly focused on health inequities.

Session 7: health inequities and racism—pathways and measurement

This session will introduce the myriad ways racial inequality can harm health and becomes embodied to create biological expressions of racism (see Table 14.1; Krieger 1999, 2012b). Relevant pathways include adverse exposure to economic and social deprivation; toxic substances, pathogens, and hazardous conditions; social trauma; targeted marketing of harmful commodities; and inadequate and degrading medical care (Krieger 1999). The session should begin by considering definitions of racism and its historical emergence and manifestations, including in relation to health, and note distinctions between—and links connecting—racism and class relations and health inequities, at the local and global levels (Banton 1998; Harrison 1999). It should then review the complexities of measuring these exposures at global, societal, institutional, community, and individual levels, in context, including actual instruments, and relating these exposures to adverse health outcomes (Krieger et al. 1993; Krieger 1999; Blank et al. 2004; Paradies 2006; Mays et al. 2007; Williams and Mohammed 2009). To make the discussion concrete, readings should also include empirical epidemiologic studies investigating links between racism and health.

Session 8: health inequities and gender—pathways and measurement

Session 8 would then focus on links between gender inequality and health: for women and men, and for girls and boys. Thinking clearly about these connections requires distinguishing between socially constructed gender and sex-linked biology (see Table 14.1; see also Doyal 1995; Krieger 2003a; Payne 2006; Fausto-Sterling 2012). It likewise calls for addressing, for any given health outcome, whether risk is affected by gender inequality, sex-linked biology, both, or neither (Krieger 2003a), as well as whether these risks are modified by other forms of social inequality (e.g. class, racism; see Krieger et al. 1993; Doyal 1995; Payne 2006). As with sessions 6 and 7, readings should include both conceptual and methodological discussions, examples of instruments, and specific epidemiologic investigations concerned with gender health inequities.

Session 9: health inequities and sexuality—pathways and measurement

The ninth session would in turn consider health inequities due to discrimination based on sexuality (see Table 14.1), most typically focused on persons engaged in consensual sex with same-sex sexual partners, and who may or may not identify as lesbian, gay, bisexual, transgender (LGBT) or other variants of sexual identity (see Parker and Gagnon 1995; Parker et al. 2004; Meyer and Northridge 2007; Committee on LGBT Health Issues 2011). Readings should include health-related review articles on conceptualizing and measuring sexuality and sexuality-based discrimination—at different levels, across the life course, and in relation to other forms of social inequality, as well as empirical studies analysing sexuality-based health inequities. While some of the epidemiologic investigations might usefully focus on HIV/AIDS and other sexually transmitted infectious (STI) diseases, for students to grasp the full health impact of inequities involving sexuality, it is critical that studies also be included on anti-LGBT discrimination and other non-STI somatic and mental health outcomes (e.g. alcohol use, tobacco-related diseases, depression, violence), and also links between sexual rights and health equity (Krieger and Sidney 1997; Parker et al. 2004; Gruskin et al. 2005; Meyer and Northridge 2007).

Session 10: health inequities between countries and regions—pathways and measurement

Session 10 would present an overview of health inequities reflecting global politics, past and present, as manifested in geopolitical—that is, country or regional—inequities in health. As with the prior sessions, readings should

include epidemiological review articles on conceptualizing and measuring global health inequities. Classroom discussion should critically analyse epidemiologic investigations, examining how the political economy of relationships between countries and regions shapes global health inequities (see Evans et al. 2001; Navarro and Muntaner 2004; Kunitz 2006; Birn et al. 2009). To aid students in grasping the magnitude of these inequities, two useful resources are (1) the 'Worldmapper' project, in which global maps scale the size of countries to the size of their health burden (Dorling et al. 2013); and (2) 'Gapminder World 2006' (Rosling 2013), which visually depicts country-level changes, from 1960 to 2004, for various health indicators, both overall and in relation to income per capita in international dollars. Other resources on global inequities in health include reports from the WHO Commission on the Social Determinants of Health (2008).

Session 11: case example—epidemiologic analyses of health inequities for a particular population

The next session would ask students to integrate their understandings of social inequalities in health by considering the cumulative embodied impact of multiple forms of inequity on a particular population, as manifested in various health outcomes (Krieger et al. 1993, 2011). The observed comorbidities may occur because either the pathogenic processes are linked (e.g. the postulated associations between metabolic syndrome and cardiovascular disease, cancer, and diabetes) or because the exposures are similarly socially patterned even if the specific etiologies are distinct (e.g. lead poisoning and cervical cancer). This could most easily be accomplished by having students critically analyse a theme issue of a journal focused on health inequities experienced by a particular population group (e.g. the 2003 issue of the *American Journal of Public Health* on racism and health among US populations of colour (Krieger 2003b)) or, if available, a theme issue or series of articles focused on one of the populations experiencing health inequities in the country of focus for the class.

Session 12: case example—epidemiologic analyses of health inequities for a particular outcome

In session 12, the perspective would be reversed, and students would be asked to consider the different types of inequities that contribute to shaping the population distribution of one selected health outcome. Assigned readings and classroom discussion could centre on a special issue devoted to health inequities involving one particular health outcome (e.g. the 2005 special issue of *Cancer Causes and Control* devoted to US cancer disparities (Krieger 2005c)), or

analogous readings for an outcome of concern in the country where the class is being taught.

Session 13: 'politically correct' or correct science? The case of racism versus 'race' and health inequities

Session 13 would then introduce students to current debates focused on links between social inequalities and health. One such debate is whether it is a matter of *correct* science (Krieger 2005d; Krieger 2007)—versus 'politically correct' science, as some influential conservative writers have charged (Satel 2000)—to conduct research on this topic. A contemporary example concerns longstanding debates over the causes of US racial/ethnic disparities in health status (Krieger et al. 1993, 1999; Williams and Mohammed 2009; Krieger 2012b). Readings should include review articles focused on the overall debate (see e.g. Krieger 2003c; Risch 2006; Williams and Mohammed 2009) and also specific case studies, to work through the in-depth meanings of this debate for epidemiologic research. Useful examples might include either (1) cardiovascular disease, contrasting etiologic research that defines the causal 'exposure' as racism (Krieger and Sidney 1996; Wyatt et al. 2003; Lewis et al. 2006) versus 'race' (Tang et al. 2006; Reiner et al. 2007); or (2) low birth weight and preterm delivery, again comparing studies investigating racism (Stancil et al. 2000; Mustillo et al. 2004; Giscombe and Lobel 2005) versus 'race' (Menon et al. 2006; DeFranco et al. 2007) as causing the observed disparities.

Session 14: 'politically correct' or correct science? Confounding and the case of hormone therapy, cardiovascular disease, and breast cancer

The example for session 14 should underscore why a concern about health inequities matters for the rigour of research not ostensibly concerned with this topic. One topical example concerns hormone therapy (HT), cardiovascular disease, breast cancer, gender, and social class (Krieger et al. 2005a). At issue is how uncritical reliance on a biomedical framework led to the discounting of epidemiologic evidence—dating back to the 1980s and recently re-confirmed—that the supposed protective effect of long-term use of HT on risk of cardiovascular disease was due to confounding by social class, reflecting how wealthier women, with better health, were the most likely to be prescribed (and could afford) HT (Petitti 2004; Lawlor et al. 2004; Krieger et al. 2005a; Rossouw 2006). This alternative hypothesis received serious attention only after the Women's Health Initiative (WHI)—the first major clinical trial to focus on HT and risk of cardiovascular disease—unexpectedly reported in 2002 that HT did not decrease, and in fact may have increased, risk of cardiovascular disease; it

also confirmed prior—albeit less well-publicized—concerns about increased risk of breast cancer (Writing Group for the Women's Health Initiative Investigators 2002). The serious potential burden of iatrogenic disease caused by HT use is shown by research indicating that the population attributable risk of breast cancer due to HT likely ranges between 10 to at least 20 per cent, which translates to an excess burden of breast cancer cases in the past decade numbering in the hundreds of thousands in the US alone (Clarke and Glaser 2007; Ravdin et al. 2007). Classroom discussion should focus on (1) the importance of addressing confounding due to the social patterning of most exposures (Davey Smith 2003; Krieger 2011), and (2) how ignoring health inequities can lead to invalid epidemiologic findings and harm the public's health (Krieger 2007).

Session 15: implications of epidemiologic research on health inequities for the public's health

Session 15 should provide examples of how epidemiologic research on health inequities can aid efforts to address these problems, from generating the evidence base to informing policy. Among possible topics of discussion, one would be current efforts to monitor the magnitude of health inequities, so that the size of the problem—and whether it is increasing or decreasing—is public knowledge; examples include the US *Public Health Disparities Geocoding Project* (Krieger et al. 2005b, 2013) and the Regional Equity Reports issued by Equinet Africa (2013). A second example could focus on the new and rapidly growing field of health impact assessment (HIA), which seeks to estimate the impact of public policies and the private sector on population health and health inequities (Krieger et al. 2003; Kemm et al. 2004; Scott-Samuel and O'Keefe 2007), including its use in new European initiatives to promote intersectoral governance for 'health in all policies' (McQueen et al. 2012). Examples of analogous epidemiologic contributions from the country on which the course is focused should likewise be included.

Session 16: who and what is accountable for social inequalities in health—summation, student projects, and course wrap-up

The final session should be used to sum up key lessons from the course; no new readings should be assigned. To ground the discussion, it would be useful to start by revisiting the initial questions posed by the course at the outset: What are social inequalities in health? And why do they matter? Students should discuss—and debate—these questions in relation to the theoretical, methodological, and empirical issues addressed in sessions 2 to 15. Time should also be

allotted for students to discuss key points learnt from doing the final assignment, and to complete a course evaluation form.

Teaching methods and format

Ideally, the course would be structured as a three-hour seminar, limited to twenty-five to thirty participants, that meets once a week for sixteen weeks. It could also be scaled down to a two-hour seminar that meets twice a week for eight weeks. If it were to be taught as a larger lecture-format course, it would need to include 'lab' sessions that give students time to discuss and debate the ideas they are learning.

The class should provide students with three opportunities to express their ideas and questions:

1. in a brief reflection-paper on each session's readings, handed in at the beginning of class (see section on 'Assessing students' achievements');
2. in a twenty- to thirty-minute small-group meeting with other students about the readings (with each group made up of six students, selected to span a range of expertise and experience), during which time the teacher would read through the reflection pieces to assess students' comprehension of the topic;
3. in a structured all-class discussion of the session's topic, led by the teacher.

To ensure a productive use of classroom time, the teacher should prepare for each class an outline of key topics to be addressed, and the amount of time allocated for discussion on each topic. The format for each class would be (1) a brief opening by the teacher in order to orient students to the session's topic, (2) small-group discussion, and (3) full-class discussion, with time left at the end for the teacher to synthesize key points raised during the class and in the readings. Time also should be provided, midway through the course, to discuss questions the students may have about the final assignment.

Assessing students' achievements

Box 14.3 describes the course's three types of assignments, intended to aid and evaluate students' achievements in fulfilling the learning objectives. The first would be the short one-page reflection papers that the students hand in at the beginning of each class (suggested as counting towards 25% of the final grade), the second would be their participation in classroom discussion (35% of the grade), and the third would be their final paper (40% of the grade). At the final session, students should complete the course evaluation form, which would be given to the teacher only after she or he has submitted the students' grades.

Box 14.3 Course format and assignments for proposed introductory class on social inequalities in health

Course format: The course is structured as a seminar, and students are responsible for participating in class discussion each session, based on the assigned readings.

Course assignments: Students will prepare for each class a short one-page reflection piece and will write a short (up to ten pages) final paper, due at the final meeting of the class.

- **The reflection piece** is intended to help organize the students' thoughts and questions before each class. It should summarize what struck the students most about the readings, what surprised them, what they learnt, and what they agreed or disagreed with and why; it should **not** simply summarize what was said in the readings. **The final paper** will critique a current epidemiologic review article on the epidemiology and etiology of a particular health outcome and which was published within the past ten years in a leading epidemiologic or public health journal (e.g. *Epidemiologic Reviews*, *Annual Review of Public Health*, *American Journal of Epidemiology*, *International Journal of Epidemiology*, etc.). The paper should

1. start with a short introduction that explains the focus and purpose of the paper (up to one page);
2. briefly describe what the article states are the key features of the population distribution of the disease and its major determinants (two to three pages);
3. critique the strengths and limitations of the article for the extent to which—and how—it discusses health inequities in relation to the outcome under consideration (four to five pages);
4. based on the materials covered in the class, offer suggestions for possible new avenues of research to identify the magnitude and causes of health inequities exhibited by the chosen outcome (one to two pages); and
5. provide a brief conclusion on whether it matters to give explicit attention to health inequities in an epidemiologic review article (up to one page).

Conclusion

Teaching about social inequalities in health is vital for epidemiology. It matters substantively and methodologically. Training students new to epidemiology in the importance of thinking rigorously about health inequities, their determinants, and their implications for epidemiologic evidence and the public's health will enhance and invigorate our field. Equipped with such knowledge, we are better positioned—in the words of Edgar Sydenstricker (1881–1935), one of the great twentieth-century social epidemiologists—to 'give glimpses of what the sanitarian has long wanted to see—a picture of the public health situation as a whole, drawn in proper perspective and painted in true colors' (Sydenstricker 1925, p. 280). With this clearer vision, we stand a better chance of producing knowledge that can make a difference in improving population health and promoting health equity.

References

Banton, M. P. (1998) *Racial Theories* (2nd edn). Cambridge: Cambridge University Press.

Berkman, L., Kawachi, I., and Glymour, M. eds. (2014) *Social Epidemiology* (2nd edn). Oxford: Oxford University Press.

Birn, A. E., Pillay, Y., and Holtz, T. H. (2009) *Textbook of International Health: Global Health in a Dynamic World* (3rd ed). New York: Oxford University Press.

Blank, R. M., Dabady, M., and Citro, C. F., eds. (2004) *Measuring Racial Discrimination*. Washington, DC: The National Academies Press.

Boucher, D., and Kelly, P., eds. (1998) *Social Justice: from Hume to Walzer*. London: Routledge.

Brandt, A. M. (2007) *The Cigarette Century: The Rise, Fall, and Deadly Persistence of the Product That Defined America*. New York: Basic Books.

Carter-Pokras, O. and Bacquet, C. (2002) What is a "health disparity"? *Public Health Reports*, **117**: 426–34.

Clarke, C. A. and Glaser, S. L. (2007) Declines in breast cancer after the WHI: apparent impact of hormone therapy. *Cancer Causes and Control*, **18**: 847–52.

Cohen, S., Kessler, R. C., and Underwood, L. (1995) *Measuring Stress: A Guide for Health and Social Scientists*. New York: Oxford University Press.

Committee on Lesbian, Gay, Bisexual, and Transgender (LGBT) Health Issues and Research Gaps and Opportunities. (2011) *The Health of Lesbian, Gay, Bisexual, and Transgender People*. Washington, DC: National Academy Press.

Davey Smith, G., ed. (2003) *Health Inequalities: Lifecourse Approaches*. Bristol: Policy Press.

DeFranco, E., Teramo, K., and Muglia, L. (2007) Genetic influences on preterm birth. *Seminars in Reproductive Medicine*, **25**: 40–51.

Diez-Roux, A. V. (2002) A glossary for multilevel analysis. *Journal of Epidemiology and Community Health*, **56**: 588–94.

Dorling, D., Newman, M., Allsopp, G., Barford, A., Wheeler, B., Pritchard, J., and Hennig, B. D. (2013) *Worldmapper: The World As You've Never Seen It Before*. <http://www.worldmapper.org/>, accessed 2 January 2013.

Doyal, L. (1979) *The Political Economy of Health*. London: Pluto Press.

Doyal, L. (1995) *What Makes Women Sick: Gender and the Political Economy of Health*. New Brunswick, NJ: Rutgers University Press.

Equinet Africa. (2013) *Regional Equity Watch 2012: Assessing Progress Towards Equity in Eastern and Southern Africa*. <http://www.equinetafrica.org/>, accessed 2 January 2013.

Evans, T., Whitehead, M., Diderichsen, F., Bhuiya, A., Wirth, M., and Whitehead, M., eds. (2001) *Challenging Inequities in Health: From Ethics to Action*. Oxford: Oxford University Press.

Fausto-Sterling, A. (2012) *Sex/gender: Biology in a Social World*. New York: Routledge

Friere, P. (1970) *Pedagogy of the Oppressed*. Tr. M. B. Ramos. New York: Seabury Press.

Gilbert, S.F. and Epel, D. (2009) *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution*. Sunderland, MA: Sinaeur Associates, Inc.

Giscombe, C. L. and Lobel, M. (2005) Explaining disproportionately high rates of adverse birth outcomes among African Americans: the impact of stress, racism, and related factors in pregnancy. *Psychological Bulletin*, **131**: 662–83.

Gordon, D., and Spicker, P., eds. (1999) *The International Glossary on Poverty*. London: Zed Books.

Graham, H. (1996) Smoking prevalence among women in the European community 1950–1990. *Social Science and Medicine*, **43**: 243–54.

Gruskin, S., Grodin, M. A., Annas, G. J, and Marks, S. P., eds. (2005) *Perspectives on Health and Human Rights* (2nd edn). New York: Routledge.

Harrison, M. (1999) *Climates and Constitutions: Health, Race, Environment and British Imperialism in India, 1600–1850*. New Delhi: Oxford University Press.

Hofrichter, R., ed. (2003) *Health and Social Justice: Politics, Ideology, and Inequity in the Distribution of Disease*. San Francisco, CA: Jossey-Bass.

International People's Health University. (2013) *Welcome to IPUH website!* <http://www.iphu.org/>, accessed 2 January 2013.

Jary, D., and Jary, J., eds. (1995) *Collins Dictionary of Sociology* (2nd edn). Glasgow: HarperCollins Publishers.

Keller, E.F. (2010) *The Mirage of a Space Between Nature and Nurture*. Durham, NC: Duke University Press.

Kemm, J., Parry, Y., and Pahner, S., eds. (2004) *Health Impact Assessment: Concepts, Theory, Techniques, and Applications*. Oxford: Oxford University Press.

Krieger, N. (1994) Epidemiology and the web of causation: has anyone seen the spider? *Social Science and Medicine*, **39**: 887–903.

Krieger, N. (1999) Embodying inequality: a review of concepts, measures, and methods for studying health consequences of discrimination. *International Journal of Health Services*, **29**: 295–352. (Republished and updated as Krieger, N. (2000) 'Discrimination and health', in L. Berkman and L. Kawachi, eds, *Social Epidemiology*. Oxford: Oxford University Press, pp. 36–75.)

Krieger, N. (2000) Epidemiology and social sciences: towards a critical reengagement in the 21st century. *Epidemiologic Reviews*, **11**: 155–63.

Krieger, N. (2001a) Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology*, **30**: 668–77.

Krieger, N. (2001b) A glossary for social epidemiology. *Journal of Epidemiology and Community Health*, **55**: 693–700.

Krieger, N. (2003a) Genders, sexes, and health: what are the connections—and why does it matter? *International Journal of Epidemiology*, **32**: 652–7.

Krieger, N., guest ed. (2003b) Theme issue: racism and health. *American Journal of Public Health*, **93**: 189–255.

Krieger, N. (2003c) Does racism harm health? Did child abuse exist before 1962?—on explicit questions, critical science, and current controversies: an ecosocial perspective. *American Journal of Public Health*, **93**: 194–9.

Krieger, N. (2005a) Defining and investigating social disparities in cancer: critical issues. *Cancer Causes and Control*, **16**: 5–14.

Krieger, N. (2005b) Embodiment: a conceptual glossary for epidemiology. *Journal of Epidemiology and Community Health*, **59**: 350–5.

Krieger, N., guest ed. (2005c) Special issues on social disparities in cancer. *Cancer Causes and Control*, **16**: 1–74.

Krieger, N. (2005d) Stormy weather: "race," gene expression, and the science of health disparities. *American Journal of Public Health*, **95**: 2155–60.

Krieger, N. (2007) Why epidemiologists cannot afford to ignore poverty. *Epidemiology*, **18**: 658–63.

Krieger, N. (2008) Proximal, distal, and the politics of causation: what's level got to do with it? *American Journal of Public Health*, **98**: 221–30.

Krieger, N. (2011) *Epidemiology and the People's Health: Theory and Context*. New York: Oxford University Press.

Krieger, N. (2012a). Who and what is a "population"? Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. *Milbank Quarterly*, **90**: 634–81.

Krieger, N. (2012b). Methods for the scientific study of discrimination and health: from societal injustice to embodied inequality – an ecosocial approach. *American Journal of Public Health*, **102**: 936–45.

Krieger, N (2013) History, biology, and health inequities: emergent embodied phenotypes and the illustrative case of the breast cancer estrogen receptor. *American Journal of Public Health*, **103**: 22–7.

Krieger, N. et al. (2005a) Hormone replacement therapy, cancer, controversies and women's health: historical, epidemiological, biological, clinical and advocacy perspectives. *Journal of Epidemiology and Community Health*, **59**: 740–8.

Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., and Subramanian, S. V. (2005b) Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *American Journal of Public Health*, **95**: 312–23.

Krieger, N., Northridge, M., Gruskin, S., Quinn, M., Kriebel, D., Smith, G. D., Bassett, M., Rehkopf, D. H., and Miller, C. (2003) Assessing health impact assessment: multidisciplinary and international perspectives. *Journal of Epidemiology and Community Health*, **57**: 659–62.

Krieger, N., Rowley, D. L., Herman, A. A., and Avery, B. (1993) Racism, sexism, and social class: implications for studies of health, disease, and well-being. *American Journal of Preventive Medicine*, **9** Suppl. 6: 82–122.

Krieger, N. and Sidney, S. (1996) Racial discrimination and blood pressure: the CARDIA study of young black and white adults. *American Journal of Public Health*, **86**: 1370–8.

Krieger, N. and Sidney, S. (1997) Prevalence and health implications of anti-gay discrimination: a study of black and white women and men in the CARDIA cohort. *International Journal of Health Services*, **27**: 157–76.

Krieger, N., Waterman, P. D., Chen, J. T., Rehkopf, D. H., and Subramanian, S. V. (2013) (As of 1 July 2004) *Geocoding and Monitoring US Socioeconomic Inequalities in Health: An Introduction to Using Area-Based Socioeconomic Measures—The Public Health Disparities Geocoding Project Monograph*. <http://www.hsph.harvard.edu/thegeocodingproject/>, accessed 2 January 2013.

Krieger, N., Williams, D., and Moss, N. (1997) Measuring social class in US public health research: concepts, methodologies and guidelines. *Annual Review of Public Health*, **18**: 341–78.

Kuh, D. and Ben-Shlomo, Y., eds. (2004) *A Life Course Approach to Chronic Disease Epidemiology* (2nd edn). Oxford: Oxford University Press.

Kunitz, S. (2006) *The Health of Populations: General Theories and Particular Realities*. Oxford: Oxford University Press.

Lawlor, D. A., Davey Smith, G., and Ebrahim, S. (2004) Socioeconomic position and hormone replacement therapy use: explaining the discrepancy in evidence from observational and randomized controlled trials. *American Journal of Public Health*, **94**: 2149–54.

Lewis, T. T., Everson-Rose, S. A., Powell, L. H., Matthews, K. A., Brown, C., Karavolos, K., Sutton-Tyrrell, K., Jacobs, E., and Wesley, D. (2006) Chronic exposure to everyday discrimination and coronary artery calcification in African-American women: the SWAN heart study. *Psychosomatic Medicine*, **68**: 362–8.

Lewontin, R. (2000) *The Triple Helix: Gene, Organism, and Environment*. Cambridge, MA: Harvard University Press.

Marmot, M. G. (2004) *The Status Syndrome: How Social Standing Affects Our Health and Longevity*. New York: Times Books/Henry Holt.

Marshall, G., ed. (1994) *The Concise Oxford dictionary of Sociology*. Oxford: Oxford University Press.

Mays, V. M., Cochran, S. D., and Barnes, N. W. (2007) Race, race-based discrimination, and health outcomes among African Americans. *Annual Review of Psychology*, **58**: 201–25.

McMichael, A. J. (2001) *Human Frontiers, Environments, and Disease: Past Patterns, Uncertain Futures*. Cambridge: Cambridge University Press.

McMichael, A. J. and Butler, C. D. (2011) Promoting global population health while constraining the environmental footprint. *Annual Review of Public Health*, **32**: 179–97.

McQueen, D. V., Wismor, M., Lin, V., Jones, C. M., and Davies, M. (2012) *International Governance for Health in All Policies: Structure, Actions, and Experiences*. Copenhagen: WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies.

Menon, R., Merialdi, M., Lombardi, S. J., and Fortunato, S. J. (2006) Differences in the placental membrane cytokine response: a possible explanation for the racial disparity in preterm birth. *American Journal of Reproductive Immunology*, **56**: 112–18.

Meyer, I. H. and Northridge, M. E., eds. (2007) *The Health of Sexual Minorities: Public Health Perspectives on Lesbian, Gay, Bisexual, and Transgender Populations*. New York: Springer.

Mustillo, S., Krieger, N., Gunderson, E. P., Sidney, S., McCreath, H., and Kief, C. I. (2004) Self-reported experiences of racial discrimination and black-white differences in preterm and low-birth weight deliveries: the CARDIA study. *American Journal of Public Health*, **94**: 2125–31.

National Association of County and City Health Officials (NACCHO). (2013) *Roots of Health Inequity*. <http://rootsofhealthinequity.org/>, accessed 2 January 2013.

Navarro, V. and Muntaner C., eds. (2004) *Political and Economic Determinants of Population Health and Well-Being: Controversies and Developments*. Amityville, NY: Baywood Publishing Company.

Oakes, J. M. and Kaufman, J. S., eds. (2006) *Methods in Social Epidemiology*. San Francisco, CA: Jossey-Bass.

Paradies, Y. (2006) A systematic review of empirical research on self-reported racism and health. *International Journal of Epidemiology*, **35**: 888–901.

Parker, R., di Mauro, D., Filiano, B., Garcia, J., Muñoz-Laboy, M., and Sember, R. (2004) Global transformation and intimate relations in the 21st century: social science research on sexuality and the emergence of sexual health and sexual right frameworks. *Annual Review of Sex Research*, **15**: 362–98.

Parker, R. G. and Gagnon, J. H., eds. (1995) Conceiving Sexuality: Approaches to Sex Research in a Post-Modern World. New York: Routledge.

Payne, S. (2006) *The Health of Men and Women*. Cambridge: Polity Press.

Petitti, D. (2004) Commentary: hormone replacement therapy and coronary heart disease: four lessons. *International Journal of Epidemiology*, **33**: 461–3.

Piermsa, T. and van Gils, J. A. (2011) *The Flexible Phenotype: A Body-Centered Integration of Ecology, Physiology, and Behavior*. New York: Oxford University Press.

Porter, D. (1999) *Health, Civilization and the State*: *A History of Public Health from Ancient to Modern Times*. London: Routledge.

Ravdin, P. M., Cronin, K. A., Howlader, N., Berg, C. D., Chlebowski, R. T., Feuer, E. J., Edwards, B. K., and Berry, D. A. (2007) The decrease in breast-cancer incidence in 2003 in the United States. *New England Journal of Medicine*, **356**: 1670–4.

Reiner, A. P., Carlson, C. S., Ziv, E., Iribarren, C., Jaquish, C. E., and Nickerson, D. A. (2007) Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African-American participants in the CARDIA study. *Human Genetics*, **121**: 565–75.

Risch, N. (2006) Dissecting racial and ethnic differences. *New England Journal of Medicine*, **354**: 408–11.

Rosling, H. (2013) *Gapminder for a Fact-Based World View*. <http://www.gapminder.org/>, accessed 2 January 2013.

Rossouw, J. E. (2006) Implications of recent clinical trials of postmenopausal hormone therapy for management of cardiovascular disease. *Annals of the New York Academy of Sciences*, **1089**: 444–53.

Satel, S. L. (2000) *PC, MD*: *How Political Correctness is Corrupting Medicine*. New York: Basic Books.

Scott-Samuel, A. and O'Keefe, E. (2007) Health impact assessment, human rights and global public policy: a critical appraisal. *Bulletin of the World Health Organization*, **85**: 212–17.

Shaw, M., Galobardes, B., Lawlor, D., Lynch, J., Wheeler, B., and Davey Smith, G. (2007) *The Handbook of Inequality and Socioeconomic Position: Concepts and Measures*. Bristol: Policy Press.

Smelser, N. J. and Baltes, P. B., eds. (2004*) International Encyclopedia of the Social and Behavioral Sciences.* <http://wwwsciencedirect.com/science/reference-works/9780080430768>, accessed 26 August 2007.

Stancil, T. R., Hertz-Picciotto, I., Schramm, M., and Watt-Morse, M. (2000) Stress and pregnancy among African American women. *Pediatric and Perinatal Epidemiology*, **14**: 127–35.

Subramanian, S. V., Jones, K., and Duncan, C. (2003) 'Multilevel methods for public health research', in I. Kawachi **and** L. F. Berkman, eds, *Neighborhoods and Health*. New York: Oxford University Press, pp. 65–111.

Susser, M. (1996) Choosing a future for epidemiology: II. From black boxes to Chinese boxes and eco-epidemiology. *American Journal of Public Health*, **86**: 674–7.

Sydenstricker, E. (1925) The incidence of illness in a general population group: general results of a morbidity study from December 1, 1921 through March 31, 1924, Hagerstown, Md. *Public Health Reports*, **40**: 279–91.

Tang, H., Jorgenson, E., Gadde, M., Kardia, S. L., Rao, D. C., Zhu, X., Schork, N. H., Hanis, C. L., and Risch, N. (2006) Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. *Human Genetics*, **119**: 624–33.

Tesh, S. N. (1988) *Hidden Arguments*: *Political Ideology and Disease Prevention Policy*. New Brunswick, NJ: Rutgers University Press.

Townsend, P. (1993) *The International Analysis of Poverty*. New York: Harvester/Wheatsheaf.

United Nations. (1948). *Universal Declaration of Human Rights*. G.A. Res 217A(III), UN GAOR, Res. 71, UN Doc A/810.

United Nations Development Programme (UNDP). (2006) *Human Development Report 2006: Beyond Scarcity: Power, Poverty, and the Global Water Crisis*. New York: UNDP.

Whiteford, L. and Whiteford, S., eds. (2005) *Globalization, Water, and Health*: *Resource Management in Times of Scarcity*. Santa Fe, NM: School of American Research Press.

Whitehead, M. (1992) The concepts and principles of equity and health. *International Journal of Health Services*, **22**: 429–45.

Williams, R. (1983) *Keywords: A Vocabulary of Culture and Society* (revd edn). New York: Oxford University Press.

Williams, D. R. and Mohammed, S. A. (2009) Discrimination and racial disparities in health: evidence and needed research. *Journal of Behavioral Medicine*, **32**: 20–47.

World Health Organization (WHO) Commission on the Social Determinants of Health (CSDH) (2008) *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health*. Geneva: WHO.

Wright, E. O. (1997) *Class Counts*: *Comparative Studies in Class Analysis*. New York: Cambridge University Press.

Writing Group for the Women's Health Initiative Investigators. (2002) Risk and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomised controlled trial. *Journal of the American Medical Association*, **288**: 321–33.

Wyatt, S. B., Williams, D. R., Calvin, R., Henderson, F. C., Walker, E. R., and Winters, K. (2003) Racism and cardiovascular disease in African Americans. *American Journal of Medical Science*, **325**: 315–31.

Young, T. K. (1998) *Population Health*: *Concepts and Methods*. New York: Oxford University Press.

Ziman, J. M. (2000) *Real Science*: *What It Is, and What It Means*. Cambridge: Cambridge University Press.

Chapter 15

Climate change and human health: issues for teacher and classroom

Ulisses Confalonieri and Shilu Tong

Introduction to climate change and human health: issues for teacher and classroom

Global climate change is the most critical and urgent issue currently facing the global society. It is a major environmental and developmental issue and constitutes a new challenge for the initiatives to protect human health (WWF 2006; Stern 2007; McMichael 2013).

Because of its newness, complexity, and scale, it poses important challenges to epidemiologists. These include some 'boundary' issues in relation to the scope, responsibility, and capacity of epidemiological research. That, in turn, presents a further challenge to the ongoing evolution of epidemiological research methods, historically driven by the ever-changing agenda of public health problems, social concerns, and prevailing theories.

In the long history of the modern human species, spanning approximately 200,000 years, global and regional climates have changed often, naturally. This has resulted in stress and, perhaps, distress to human communities and cultures (McMichael 2012). The collapse of the Mayan civilization in Central America around 1,100 years ago and that of the smaller West Viking civilization in southwestern Greenland around 700 years ago were each attributable to natural climatic changes occurring over the course of a century or two—the former due to a long-term drying cycle, and the latter to a progressive cooling that led into the Little Ice Age in Europe (McMichael 2001; Diamond 2005). Today we face, for the first time, *human-induced* climate change—and it is occurring at a rate that is an order of magnitude faster than changes that occurred previously from natural sources.

Climate scientists foresee unavoidable further climate change over at least the coming half-century because of the accrued emissions of greenhouse gases in the atmosphere with their as yet unrealized consequent warming, topped up by ongoing emissions.

Climate change will affect many of the familiar, proximal determinants of health. These include levels of thermal stress, food availability, and nutrient quality, the range and activity of infectious diseases, the occurrence of extreme climatic/environmental events (e.g. floods, fires, storms, and droughts), and the viability of livelihoods (especially agricultural).

A good overview of the issues pertaining to health risks and their reduction can be obtained from chapter 8 of the *Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Confalonieri, Menne, et al. 2007; available online at <http://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4-wg2-chapter8.pdf>) and also from the special report of the Intergovernmental Panel on Climate Change (IPCC) on extreme events (Field et al. 2012; available online at <http://ipcc-wg2.gov/SREX/report/>).

Teaching objectives

Teaching this topic to students of epidemiology entails both conceptual novelty and various methodological challenges. The topic also has great, even urgent, social relevance to the evolving conditions of human life in today's world.

Familiar epidemiological methods are directly applicable to many aspects of this topic. They are useful for studying how variations and trends in climatic conditions affect health outcome rates. Indeed, there remain many gaps in that essential information base. The concept of population-attributable risk is directly applicable to estimating current disease burdens attributable to extant climate change. Epidemiological methods, both non-experimental and experimental, are applicable to the evaluation of adaptive strategies to lessen the risks to health from climate change and its environmental and social sequelae.

In formulating a teaching curriculum, there are several distinctive features that warrant attention. These include the following:

- Understanding the nature and scale of this human-induced environmental change. This requires familiarity with various new concepts and technical terms, and with literature from often unfamiliar disciplinary areas.
- Learning about the nature of complex systems-based changes in the natural environment. 'Climate' is not a single simple exposure agent, acting via direct insult to human biology. Some pathways by which climate change influences health are indirect and multistage, including via changes to complex non-linear natural environmental systems and social systems.
- The need for an appreciation of the sources and types of uncertainties that are inherent in much of this topic, how they can be quantified, and how best to communicate them to end-users.

- Understanding the diversity of epidemiological research methods needed in this complex topic domain. These span empirical studies (at various scales) of risk identification and quantification, scenario-based modelling of future health risks, and the optimization (e.g. equity, cost, timeliness, etc.) and evaluation of preventive (adaptive) strategies to reduce risks to health.
- Learning about data-analytic techniques appropriate to this domain of research, with particular emphasis on the analysis of population-level data. This includes time-series analysis, spatial analyses (geographic information systems (GIS), etc.), the acquisition and use of remote-sense data, and scenario-based modelling of future health risks from climate change.

This is a relatively new, and now rapidly evolving, research area for which there is as yet limited experience in curriculum design and teaching methods. We are therefore all pioneers in undertaking this challenging task of building research capacity applicable to climate change.

Teaching content

Basic climate science

The basic science of climate change is best learnt from reports of the UN's IPCC (<http://www.ipcc.ch/index.htm>). The main resource is the five-yearly sequence of three-volume Assessment Reports (1991, 1996, 2001, 2007, 2013). Each volume has a 'Summary for policy makers'—and that of Working Group I, the *Fifth Assessment Report* (IPCC 2013), contains a succinct, referenced, well-illustrated text about current understanding of the climate change process. Some of the main conclusions from Working Group I and also from the IPCC special report on extreme events (Field et al. 2012) are summarized in Box 15.1.

Issues of scale: the spectrum of epidemiological research tasks

Climate change is, at first sight, environmental health 'writ large'. Where urban air pollution impinges on whole local communities, a change in climatic conditions can affect the severity of heat waves over continental Europe, the food yields of an entire region such as Sub-Saharan Africa, or the geographic range of transmission of dengue fever in the Americas. Therefore, many important research questions refer to understanding the health risks faced by whole populations or regions (Patz et al. 2005). Nevertheless, many other important epidemiological research questions exist at more disaggregated levels. Local communities, subgroups, and categories of persons may all display differences insusceptibility to climate-related health risks—for example, to heat wave impacts, nutrient deficits or food shortages, or an increased incidence of infectious disease.

Box 15.1 Scientific assessment of observed climate change by the Intergovernmental Panel on Climate Change

The Intergovernmental Panel on Climate Change (IPCC) concluded that

- the average temperature of the Earth's surface has risen by about 0.7°C since 1900;
- the eleven warmest years on record since 1850 have occurred in the past twelve years;
- the average sea level has risen 170 mm since 1900 and has risen at an increased rate of 3 mm per year since 1993;
- there is a projected increase in duration and intensity of droughts in some regions;
- there is an observed increase in the length or number of warm spells or heat waves in many regions;
- there will likely be increases in the frequency of heavy precipitation and in rainfall totals from heavy precipitation over many areas of the globe;
- the water vapour content of the atmosphere has increased since at least 1980, consistent with the theory that warmer air can hold more moisture (water vapour is itself an important greenhouse gas); and
- the oceans have become more acidic due to higher concentrations of carbon dioxide.

The IPCC also concluded that there is a high likelihood that human-induced increases in greenhouse gases have caused most of the observed increase in globally averaged temperatures since the mid-twentieth century. This important conclusion was based on a range of specific types of geophysical and meteorological evidence.

Source: Data from *The Intergovernmental Panel on Climate Change (IPCC)*, Copyright © IPCC 2014, available from http://www.ipcc.ch/index.htm

The health risks from climate change also pose many research challenges appropriate to the concepts and methods of social-environmental epidemiology. When economic and social stresses arise as a result of climate change, and when vulnerable (coastal, food-insecure, water-deprived, etc.) populations are displaced and their social structures and livelihoods disrupted, many types of physical and mental health risks are likely to occur.

For the above reasons, much of the epidemiological research appropriate to studying and reducing the health risks from climate change requires the analysis of population-level indices of exposure and health outcome. Besides, as for environmental epidemiology in general, many climate-related exposures are difficult to measure at the localized level, let alone at the individual level. In seeking risk-lessening strategies, analysis and intervention at the 'population level' will often lead to the most effective, long-term, intervention is possible—a point that echoes Geoffrey Rose's important dictum that epidemiologists should pay as much attention to explaining disease *rates* as to explaining disease *cases*. Indeed, Rose argued that effective preventive strategies very often require that we understand what makes whole populations sick (manifested by an increased rate of disease or pre-disease X), and not just what makes particular individuals within any such population sick (Rose 1985).

This question, rephrased, is always an important general issue for classroom discussion: 'To what extent is the likely level of effective health-protecting intervention by society also the level at which epidemiologists should focus their research into "causes"?' The many and diverse studies of air pollution and health, conducted over the past two to three decades and particularly in the US and Europe, provide a good starting point for this discussion.

The need for higher-resolution (disaggregating) epidemiological research applies to issues of space, time, and population type. Studies of geographically localized effects of climate change are often needed. For example, a recent study on the north Alaskan coast revealed that the coastal water temperature had risen over the past decade and, by around 2004, had reached the critical summertime level of approximately 15°C, which allowed sustained bacterial contamination of oyster beds and thus posed diarrheal food-poisoning risks to local communities and seasonal tourist parties (McLaughlin et al. 2005). Studies in Paris after the notorious 2003 heat wave that killed several tens of thousands of Europeans showed that the risks were particularly high for elderly persons in retirement homes. This echoed an earlier finding that most of the deaths during the extreme heat wave in Chicago in 1995 were in inner-urban, poor, black families living in unventilated apartment housing (Semenza et al. 1996).

There is, clearly, a challenging and urgent range of epidemiological research to be undertaken to identify, quantify, and lessen the health risks posed by climate change. There are roles here for environmental epidemiologists, mathematical modellers, and social epidemiologists. The latter investigate the social determinants of population patterns of health and disease, rather than treating such determinants as mere background to biomedical phenomena (Krieger 2001a). Research inputs from the social sciences and humanities can thus

extend the scope of research to better understand the roles of social and economic structures, institutions, and human behaviours in (1) the underlying causation of climate change, (2) determining the vulnerability of specific exposed communities, and (3) shaping the social and political responses to both the environmental hazard and the health risks.

Risks to human health

Climate change will have many adverse health impacts. Some benefits to health may also occur in some locations—although they may not persist in future under conditions of further climate change. For example, malaria may recede in some low-latitude regions (e.g. parts of West Africa, the Chittagong Hills in Bangladesh, etc.) because temperatures and/or drying may increase to levels that preclude mosquito survival.

The three main causal pathways are summarized in Fig. 15.1: (1) direct stresses on human biology, (2) indirect pathways that entail changes in physical and ecological systems and processes, and (3) socially mediated risks following environmental disruption of livelihoods and community resources.

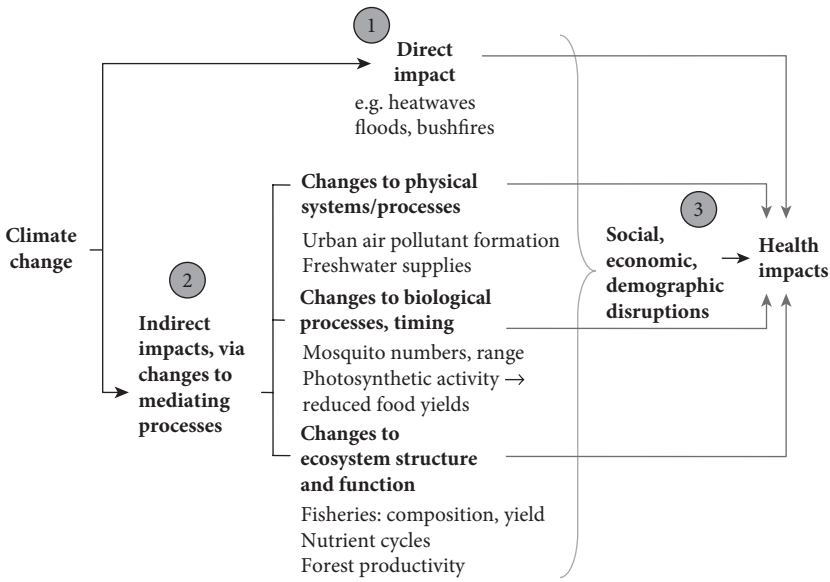


Fig. 15.1 The main pathways by which climate change affects human health. Note the progression—from category 1 (relatively simple and direct) through category 2 (changes to complex environmental and ecological systems) to category 3 (health risks from the social–economic–demographic consequences of climatic–environmental disturbances).

The detection and quantification of these health risks pose challenges. The melting of glaciers and many of the changes in *non-human* biological systems—germinating, flowering, nesting, feeding, and migration timing—are readily attributable to changes in temperature. Human populations, by contrast, are often well insulated against climatic stresses via culture (housing, clothing, etc.) and associated behaviours. Hence, the attribution of human health impacts to climate change is less easy. If malaria ascends to higher altitude in eastern Africa, is it due to changes inland use, population movement, the cessation of mosquito-control programmes, the emergence of anti-malarial resistance, or regional warming—or all of the above?

Recent scientific papers suggest an emerging influence of climate change on infectious disease patterns. This includes the northwards extension of the vector tick for tick-borne encephalitis in Sweden (Lindgren and Gustavson 2001; Lindgren and Jaenson 2006) associated with warming winters, the northwards extension of the water snail that spreads schistosomiasis in eastern China and some evidence of an increased range of the disease itself (Yang et al. 2005; Zhou et al. 2008), and the ascent of highland malaria to higher altitudes in parts of eastern and southern Africa in association with local warming (e.g. Pascual et al. 2006; Chaves and Koenraadt 2010; Alonso et al. 2011). Each of these and other such examples, on their own, remains inconclusive at this relatively early stage, since it is difficult to assess the extent to which climate change is responsible for disease shifts relative to other coexistent (confounding) influences. Viewed in aggregate, however, the evidence is becoming persuasive.

Climate-related environmental refugees

This aspect of the topic poses a boundary issue. Epidemiology is the central, quantitative, science of public health. So, how should we tackle the potentially great but difficult-to-measure health consequences of social and economic disruption and the consequent population displacement?

If climatic and environmental conditions deteriorate in vulnerable regions, there will be increased pressures from environmental refugees on neighbouring (especially richer) countries. The demographic disruption and associated social tensions will be accompanied by adverse health consequences that typically affect displaced persons—and, perhaps, their hosts (see also pathway 3 in Fig. 15.1). This includes an increased risk of mental health problems, infectious diseases, and malnutrition (McMichael 2006); see, for example, the analysis of how regional climate change in the Sudan and Chad region of Africa has contributed to the Darfur conflict (UNEP 2007).

Discussing the role of 'ecological' studies—the population perspective

In classroom discussion of this topic, one particular question is likely to arise as a result of the received textbook wisdom that 'ecological' studies are poor cousins of good epidemiological research. This (regrettably inflexible) orthodoxy states that population-level analysis can describe relationships and may generate hypotheses but is severely limited as a tool for discovering and quantitating risks to health. Further, there may be 'ecological fallacies' at play, resulting in inappropriate cross-level inference (Pearce 2000).

For many etiological relationships, discoverable and alterable at individual level, that critique is indeed fair comment. Most of our knowledge of the health risks due to dietary habits, smoking, alcohol consumption, sex-hormone use, and physical activity has come from cohort and case-control studies, as has much of our understanding of occupational health hazards. For many other etiological relationships, for which whole communities or populations are the natural unit of exposure, that view is ill-informed and a barrier to good research. Among very many important historical examples, consider the study designs used to discover the beneficial health effect of fluoride in drinking-water or the mortality impact of heat waves.

There are, therefore, no final 'correct' answers here as to the legitimacy of specific methods. However, which is cart and which is horse? Should our existing repertoire of research methods set boundaries on what we study? Or should epidemiologists take their cue from the population health concerns of society-at-large and develop methods that give the best possible answers to those health issues that press most on the population that employs us?

The latter path has seen epidemiological research methods progress from community-level 'miasmatic' (i.e. assuming pervasive environmental exposures) and sociological investigations of the disease and mortality differentials of early industrialization, to the higher-resolution studies of specific occupational health risks (radiation, lead exposure, aromatic amine exposures, asbestos inhalation, etc.) of the late nineteenth and early twentieth centuries. The rise of the germ theory, which refocused the risks of infectious disease at individual level, and the subsequent new insights into micronutrient deficiencies early last century redirected epidemiologists to addressing disease risks at the more local, family, and individual levels.

The resultant strengthening of the biomedical model, focused on individual-level exposures and biological responses, provided a ready base for studying the rise of non-infectious diseases in late-industrial societies from around the mid-twentieth century. This underpinned the rise of modern epidemiology, with its

very successful methods of individual-level study, especially cohort and case-control study designs.

Yet something was missing from this research agenda. Epidemiologists were failing to study and elucidate many population-level health issues—such as the unequal distribution of health risks within populations, the patterns of emergence and spread of various infectious diseases, and why coronary heart disease rates had begun to fall in high-income countries. In response to this deficit, the above-mentioned field of social epidemiology has strengthened recently. It seeks in particular to understand the dynamics of upstream, or distal, influences on patterns of health and health-related behaviours in populations.

Within that enlarged research frame, other multilevel ecologically oriented approaches are emerging that help us integrate social and biological reasoning and thus elucidate the determinants of population health (Krieger 2001b). These approaches, including 'ecosocial theory' (Krieger 1994), 'eco-epidemiology' (Susser and Susser 1996), the 'social-ecological systems perspective' (McMichael 1999), and the 'ecosystem approach to health' (Forget and Lebel 2001), all invoke ecological models that seek to understand how humans interact with their social and natural environments (Krieger 2001b).

Today, as 'globalization' proceeds, as populations and countries around the world become more interconnected, and as the scale of human demand on the natural environment increases, so there is need to frame some of our epidemiological research questions at that level. Climate change and its many environmental and social ramifications represent a major, urgent area of need for this type of innovative epidemiological research.

Climate change and health: a key issue for 'sustainable development'

The advent of global climate change broadens the research agenda for epidemiologists. It is interesting that, in the 1990 Global Burden of Disease (GBD) assessment, the World Health Organization (WHO) did not include climate change as a risk factor (indeed, only a few epidemiologists anywhere had at that time given it any real thought). For the 2000 GBD assessment, however, climate change was recognized as important. This issue has now been mainstreamed, as it becomes increasingly apparent that climate change is affecting health risks around the world.

Students should be asked to discuss the following question: 'Do we really understand that the ultimate determination of population health, over decadal time, lies in the conditions of the natural and social environments? Do our prevailing theories of disease causation and our dominant research methods take

account of this fundamental reality? Or is that beyond the boundaries of epidemiology?'

This requires us to stand back from the detail in the well-studied foreground (comprising mostly individual-level epidemiological analyses of inter-individual differences in health risks) and recognize the larger, obligatory, ecological frame within which human population health is determined (note: ecological analysis, in its *true* sense, focuses on ecological relations and behaviours within a systems framework, examining interactions between communities and their environment).

Student appreciation of the character of this research topic will be enhanced by a comparison of the two kinds of environmental influence on health—influences that arise at very different scales:

1. The local, ambient, environment poses health risks via chemical, physical, or microbiological hazards. These are mostly human-generated contaminants and they do direct harm to human organ systems. Over the past quarter century, there has been much epidemiological research into this category of environmental health hazards.
2. At much larger scale are the biosphere's geophysical systems and ecosystems. These provide environmental stabilization, replenishment, recycling, and production—that is, they furnish the life-supporting processes that underpin the well-being and health of human populations. Today, the global human demand on the natural environment is causing unprecedented global environmental changes, and, hence, fundamental risks to human population health. The impacts of climate change on agricultural yields, marine productivity, freshwater supplies, natural constraints on infectious agents, and the magnitude and geography of extreme weather events all pose risks to human safety, health, and survival.

The increasing societal concern over the anticipated health impacts of human-induced global climate change poses a challenge to epidemiologists: to elucidate the risks to health, to identify vulnerable communities or populations, and to formulate and evaluate preventive and protective actions. There is also the need to develop practical methods for quantifying and comparing the vulnerability of different population groups to the health impacts of climate change, in order to optimize the choice of adaptive strategies (Thornton et al. 2008; Confalonieri et al. 2009 2014). True primary prevention, of course, requires a radical worldwide reduction in greenhouse gas emissions. Meanwhile, because climate change is already occurring and additional change is 'committed' within the climate system, no matter what action humans take today, health-protective action via 'adaptive' strategies is needed (note also the emphasis on this aspect

in the IPCC *Fourth Assessment Report*—specifically the report of Working Group II, available at: <http://www.ipcc-wg2.org/>).

A further stimulus to epidemiological research on climate change is that gains in population health are central to the achievement of the UN's 'Millennium Development Goals', and the adverse impacts of climate change are likely to undermine achieving those development goals. Climate change associated events can reduce the resilience of communities—especially of the poorest—and overwhelm the coping capacity of many societies (Haines and Cassel 2004; Confalonieri et al. 2009, 2014).

This raises for discussion again the topic of the 'boundaries' of epidemiological research. How are they to be determined? By formal definition of the discipline, by the repertoire of existing research methods, by the changing demands of society as it encounters new health hazards, or by some other means?

Curriculum development: contextual issues and perspectives

A teaching course, module, or short course should include, up front, the following four contextual elements:

1. A review of the basic science of climate change. Students should have a clear understanding of how human action is affecting the world's climate. Key resources are
   - IPCC reports, especially the synthesis report of the IPCC *Fourth Assessment Report* (see the IPCC website: <http://www.ipcc.org/>); and
   - documents containing basic information about complex, dynamic, non-linear systems. These are central to Ecosystem science—see the Millennium Ecosystem Assessment (2005).
2. Understanding the role, in forecasting future risks, of scenarios of greenhouse gas emissions: how they are specified and how they are incorporated into modelling future climatic changes (see Box 15.2). A good introduction is available in the 'Summary for policy makers' in the Working Group I volume of the IPCC *Fourth Assessment Report* (IPCC 2007)—see the IPCC website (<http://www.ipcc.org/>). Particular attention should be paid to understanding
   - the development and use of global climate models (GCMs) to model future changes in global, regional, and local climatic conditions (see also Box 15.2);
   - the sources of variation in future greenhouse gas emissions scenarios—that is, the prime input to GCM modelling of future climate change; and
   - the significance of the anticipated future changes in mean climatic conditions versus changes in climatic variability.

Box 15.2 Global climate models: a brief note

These models have been developed over the past several decades by groups of climate scientists in many locations around the world. This has occurred particularly in response to the prospect of climate change. In particular, in the past half-decade there has been progress in the understanding of important climate processes and their representation in global climate models (GCMs).

The models incorporate mathematical representations of relationships between key parameters in the responses of (and interactions between) atmosphere, oceans, terrestrium, biota, and cryosphere (glaciers, sheets) to a change in greenhouse gas concentrations and hence in the energy profile of the climate system. Today's GCMs still have some limitations (e.g. in handling water vapour and cloud cover) but climate scientists have a much increased confidence in their capacity to project future climate, particularly in light of the models' ability to simulate, on the basis of specified greenhouse gas concentrations, the following:

- current average climate;
- year-to-year variability of climate;
- extreme events, such as heat waves, floods, and storms;
- climates from thousands of years ago; and
- recently documented trends in global and regional climate.

The current generation of GCMs can credibly simulate climate conditions at global and continental scales for most meteorological variables of interest. While there is more development research to be done, downscaling to higher resolution can also provide useful insights into climate change at smaller scales.

3. Recognition that climate change does not exist in isolation. It is one of a set of human-induced 'global environmental changes' (GECs). Some of the main GECs and their impacts on some health risks are shown in Fig. 15.2. The figure also shows that many health impacts will be affected by concurrent, often interacting, impacts from several environmental changes. An illustrative example of these interactions is the set of observed ecosystem changes (species composition, reproductive patterns, migratory patterns, etc.) caused by climate change (Daszak et al. 2000; Walther et al. 2002; Parmesan and Yohe 2003; Ballard et al. 2012). These can affect the distribution and population

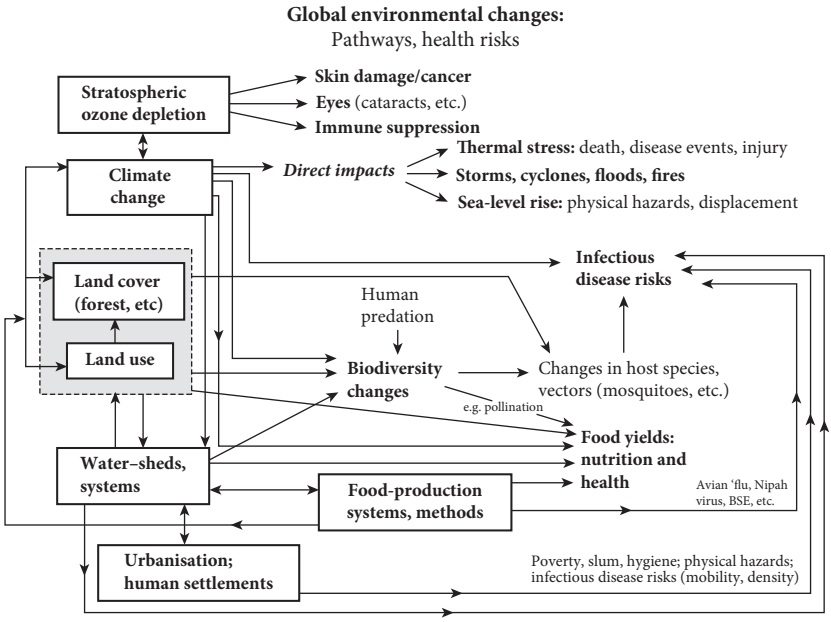


Fig. 15.2 Pathways by which the main types of global environmental change affect (selected) human health outcomes.

dynamics of different components of the cycles of zoonotic infectious diseases in various non-human hosts (e.g. leishmaniasis, Lyme disease, tick-borne encephalitis, etc.). Methods for ecological niche modelling may be used to explore the future distribution of vectors and wildlife hosts of pathogens under a changing climate. These would point either to an expansion/contraction of the disease distribution in space or to a situation of stability (Holt et al. 2009; Peterson 2009; González et al. 2010; Daszak et al. 2013).

4. Recognition of the limitations of scenario-based modelling of future health risks, especially those of vector-borne diseases, on a global scale. Besides the uncertainties inherent to the modelling of climatic scenarios, several non-climatic drivers of endemic infections, especially in tropical areas (e.g. malaria, dengue fever, and leishmaniasis) are difficult to quantify, although they play a very important role in the dynamics of these diseases. Among them are the reliability of surveillance and disease control activities and also the environmental changes that may limit the species range of important vector species. In many cases, extensive areas are currently free from a given disease due to these factors and not to climatic constraints.

Discussing the research tasks for epidemiology

Figure 15.3 displays the main 'branches' of epidemiological research activity in relation to climate change and within the framework of health-risk identification, assessment, and management. This diagram warrants detailed classroom discussion. It indicates a number of loci where epidemiological research is needed. Overall, six main categories of research are relevant, each of them requiring continuing development of epidemiological methods. The six main categories of research tasks are

1. learn from the recent past—clarify relationships between (background/natural) climate variation and health outcomes;
2. seek evidence of any actual current health impacts of climate change;
3. estimate, statistically, the current burden of disease (e.g. annual deaths, disability-adjusted life years (DALYs)) attributable to climate change;
4. develop scenario-based modelling to forecast future risk (including dealing with complexity and uncertainty);
5. estimate health co-benefits of actions taken to avert/reduce further environmental change; and
6. formulate and evaluate health-protecting ('adaptive') actions.

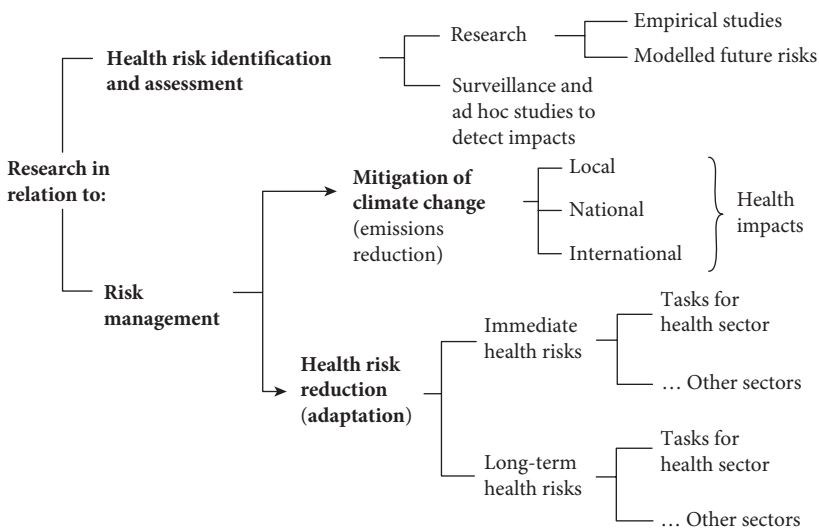


Fig. 15.3 Main components and relations of climate change health-risk identification, assessment (quantification), and management.

The first four of these research tasks refer to the 'core' area of understanding, quantifying, statistically estimating, and future-modelling the risks to health from climate change. The final two refer, respectively, to using existing epidemiological knowledge to estimate the health impacts (hopefully, health gains—i.e. 'co-benefits') of actions taken by society to mitigate climate change, and to estimating the anticipated and achieved health gains from adaptive strategies to lessen the adverse health impacts of climate change.

For the first two tasks, the main empirical approaches to investigating the climate–health linkages are

- spatial studies, where climate is an explanatory variable in the geographical distribution of the disease (Craig et al. 1999; Lindgren and Naucke 2006); and
- temporal studies, assessing the health effects of climate variability—changes in temperature or rainfall—at short time-scales (daily, weekly, seasonal), or longer-term (inter-annual, decadal) changes (Chaves and Pascual 2006; Wu et al. 2007).

Recommended references for the above research tasks

General reviews of epidemiological research: substrate for class discussion

- McMichael, A J., Woodruff, R. E., and Hales, S. (2006) Climate change and human health: present and future risks. *Lancet*, **367**: 859–69.
- Confalonieri, U., Menne, B., Akhtar, R., Ebi, K. L., Hauengue, M., Kovats, R. S., Revich, B., and Woodward, A. (2007) 'Human health', in M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, eds, *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. (See IPCC Working Group II website at <http://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4-wg2-chapter8.pdf>.)
- Frumkin, H., McMichael, A. J., and Hess, J. J. (eds). (2008) Special issue: climate change and the health of the public. *American Journal of Preventive Medicine*, 35.
- Stanke, C., Kerac, M., Proudhomme, C., Medlock, J., and Murray, V. (2013) Health effects of drought: a systematic review of the evidence. *PLoS Currents*, 5: ecurrents.dis.7a2cee9e980f91ad7697b570bcc4b004.
- Berry, H.L., K. Bowen, and Kjellstrom, T. (2010) Climate change and mental health: a causal pathways framework. *International Journal of Public Health*, **55**: 123–32.

Task 1: learning about climate–health relationships from the recent past

- Vandentorren, S., Suzan, F., Medina, S., Pascal, M., Maulpoix, A., Cohen, J. C., and Ledrans, M. (2004) Mortality in 13 French cities during the August 2003 heat wave. *American Journal of Public Health*, **94**:1518–20. Apparent 'threshold' temperature reached; perhaps the critical threshold duration of heat wave was also reached. Which subgroups were at most risk?
- Chinga-Alayo, E., Huarcaya, E., Nasarre, C., Del Aguila, R., and Llanos-Cuentas, A. (2004) The influence of climate factors on the epidemiology of bartonellosis in Ancash, Peru. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **298**: 116–24.
- Earnest, A., Tan, S. B., and Wilder-Smith, A. (2012) Meteorological factors and E Niño-Southern oscillation are independently associated with dengue infections. *Epidemiology and Infection*, **140**: 1244–51.
- Bandyopadhyay, S., Kanji, S., and Wang, L. (2012). The impact of rainfall and temperature variation on diarrheal prevalence in Sub-Saharan Africa. *Applied Geography*, **33**: 63–72.

Task 2: detection of current health impacts

- Lindgren, E. and Gustafson, R. (2001) Tick-borne encephalitis in Sweden and climate change. *Lancet*, **358**: 16–18. Did the warmer winters during the 1980s and 1990s cause the tick vector and hence the human disease to extend north?
- Zhou, X.-N. et al. (2008) Potential impact of climate change on schistosomiasis transmission in China. *American Journal of Tropical Medicine and Hygiene*, **78**: 188–94.
- Alonso, D., Bouma, M. J., and Pascual, M. (2011) Epidemic malaria and warmer temperatures in recent decades in an east African highland. *Proceedings of the Royal Society B: Biological Sciences*, **278**: 1661–9.

Task 3: estimating current burden of disease attributable to climate change

- McMichael, A. J. et al. (2004) 'Climate change', in M. Ezzati, A. D. Lopez, A. Rodgers, and C. J. L. Murray, eds, *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Due To Selected Major Risk Factors*. Geneva: World Health Organization, 1543–650, available at <http://www.who.int/publications/cra/chapters/volume2/1543–1650.pdf/>.
- Lim S. S. et al. (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions,

1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**: 2224–60.

Task 4: modelling future health risks

- Ebi, K. L., Hartman, J., Chan, N., McConnell, J., Schlesinger, M., and Weyant, J. (2005) Climate suitability for stable malaria transmission in Zimbabwe under different climate change scenarios. *Climatic Change*, **73**: 375–93. Modelling of future changes in malaria transmissibility, using fuzzy logic.
- Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A. P., and Palmer, T. N. (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**: 576–9. The authors derived, from recorded time-series data, the statistical relationship between summer (December to January) rainfall and the ensuing post-summer annual malaria incidence. They then applied this model (with reasonable success) to the recent historical record of meteorologically modelled forecasts of impending summer rainfall and demonstrated the possibility of extending by several months the summer rainfall-based warning of subsequent malaria risk.
- Huang, C., Barnett, A., Wang, X., and Tong, S. L. (2012b) The impact of temperature on years of life lost in Brisbane, Australia. *Nature Climate Change*, **2**:265–70.
- Mangal, T. D., Paterson, S., and Fenton, A. (2008) Predicting the impact of long-term temperature changes on the epidemiology and control of schistosomiasis: a mechanistic model. *PLoS One*, **3**: e1438.
- Béguin, A., Hales, S., Rocklöv, J., Åström, C., Louis, V. R., and Sauerborn, R. (2011) The opposing effects of climate change and socio-economic development on the global distribution of malaria. *Global Environmental Change*, **21**: 1209–14.

Task 5: co-benefits of mitigation actions

- Friel, S., Dangour, A. D., Garnett, T., Lock, K., Chalabi, Z., Roberts, I., Butler, A., Butler, C. D., Waage, J., McMichael, A. J., and Haines, A. (2009) Public health benefits of strategies to reduce greenhouse-gas emissions: Food and agriculture. *Lancet*, **374**: 2016–25.
- Haines, A. et al. (2009) Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers. *Lancet*, **374**: 2104–14.

- Anenberg, S. C. et al. (2012) Global air quality and health co-benefits of mitigating near-term climate change through methane and black carbon emission controls. *Environmental Health Perspectives*, **120**: 831–9.

Task 6: evaluation of adaptive strategies

- Ebi, K. L. and Schmier, J. K. (2005) A stitch in time: improving public health early warning systems for extreme weather events. *Epidemiologic Reviews*, **27**: 115–21.
- Tan, J., Zheng, Y., Song, G., Kalkstein, L. S., Kalkstein, A. J., and Tang, X. (2007) Heat wave impacts on mortality in Shanghai, 1998 and 2003. *International Journal of Biometeorology*, **51**: 193–200.
- Dodman, D., Mitlin, D., and Co, J. R. (2010) Victims to victors, disasters to opportunities: community-driven responses to climate change in the Philippines. *International Development Planning Review*, **32**: 1–26.
- Keim, M.E. (2008) Building human resilience: the role of public health preparedness and response as an adaptation to climate change. *American Journal of Preventive Medicine*, **35**: 508–16.

Health impacts of climate change: what types of epidemiological study?

Of the six categories of epidemiological research introduced in 'Discussing the research tasks for epidemiology', the first four form a 'core' that addresses aspects of the central question: what risks does climate change pose to human health, either currently or in the foreseeable future? Much of this research needs to be conducted in collaboration with other disciplines with expert knowledge in this complex field. These four core categories are

1. studying the 'baseline' relationship between climatic conditions (natural trends, variability) and specified health outcomes—this enriches the information base needed for estimating the health impacts of projected future changes in climatic conditions;
2. seeking evidence of changes in rates of occurrence of disease or disease risk factors (e.g. mosquito population range) that are reasonably attributable to local recent climate change;
3. statistically estimating, at appropriate geographic scale, the proportion of each particular known climate-sensitive health outcome reasonably attributable to the observed accompanying climate change; and
4. using information from category 1 to estimate how future, geographically gridded scenarios of climate change (generated from expectations of future

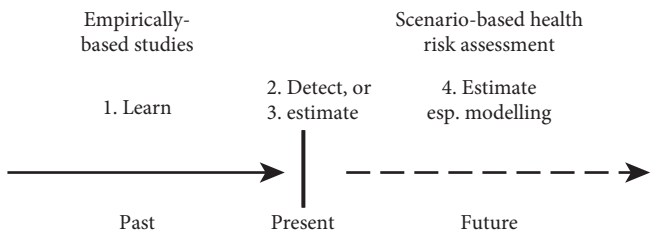


Fig. 15.4 Four main research categories: study baseline climate–health relationships, detect impacts, attribute burden of disease, and predict future health risks. Two other research tasks, not shown here, relate to the assessment of adaptive strategies and to studies of the health consequences of mitigation (emissions reduction) strategies.

greenhouse gas concentrations in the lower atmosphere) will affect the rate and/or range of occurrence of specified health outcomes.

These categories of research can be visualized across a time axis, as shown in Fig. 15.4, and are discussed in more detail below.

Clarifying basic climate–health relationships: learning from the recent past

Because of the relatively low priority accorded to studying the relationship between natural climate variability, weather events, and human health prior to the 1990s, the advent of climate change as a focus for research has necessitated much additional basic epidemiological research to fill in the missing information. Time-series studies have been very useful, as have the impacts of extreme weather events.

To assess the near-term (immediate) health effects of climate variables and air pollution, several methods of time-series analysis have been widely used. These include, in particular, generalized linear models (GLMs) with parametric splines (such as natural cubic spines) and generalized additive models (GAMS) with non-parametric smoothing splines or lowness smoothers (McCullagh and Nelder 1989; Hastie and Tibshirani 1990).

The relationships among climate conditions, air pollutants, and mortality often involve high-order interactions and multiple co-linearities. These are often difficult to handle via the above-mentioned models. In such situations, classification and regression tree (CART) analysis can provide an alternative non-parametric approach, without requiring assumptions of linear relationships among variables or homoscedasticity in variances. Using a sequence of binary splits, CART analysis segments the data into homogeneous subgroups, a method that is well suited to both to exploring and modelling such data (Breiman et al. 1984).

Detecting actual health impacts of climate change

The detection of actual health impacts of climate change at this relatively early stage of the process is difficult. The average surface temperature of the Earth has increased by around 0.6°C since the mid-1970s (most of this increase attributable to human actions) and we can be certain that there have been impacts on some health risks and outcomes in some places.

The IPCC Working Group II has concluded from its review of many thousands of reported studies that there is compelling evidence that documented regional warming has already affected many cycles and processes in nature (IPCC 2007). However, as mentioned earlier, the human species is different! Its populations are buffered by culture and technology and, anyway, climate–health causal relationships are often clouded by associated confounders. The most suggestive evidence to date of impacts is from studies of recent changes in the patterns of infectious diseases such as tick-borne encephalitis, Lyme disease, malaria, schistosomiasis, and some others.

Research challenges include

- the rational identification of likely places/circumstances to seek evidence of climate change-related alterations;
- the critical assessment of alternative explanations for observed changes in health outcomes; and
- the development of analytic techniques for teasing out shifts in temporal and spatial patterns.

Estimating current burdens of disease due to climate change

Accruing evidence from around the world shows that many systems are already being affected by recent climate change (IPCC 2007). The health impacts of climate change range from direct to indirect effects, such as extra deaths due to heat waves, increases in the transmission of climate-sensitive infectious diseases, and mental health problems caused by income loss due to a reduction in agricultural productivity associated with extreme weather events (Haines et al. 2006; McMichael 2013). It is important to estimate current burdens of disease attributable to climate change and to project future health effects under climate change scenarios.

Improved understanding of current climate–health relationships is essential to all these objectives. Both qualitative and quantitative research activities play important roles in the establishment of the baseline climate–health relationships (Fig. 15.5). Qualitative research often applies an analogous approach to describe the possible health impacts of climate change using historical evidence. For example, McMichael (2012) elegantly categorized the health risks of

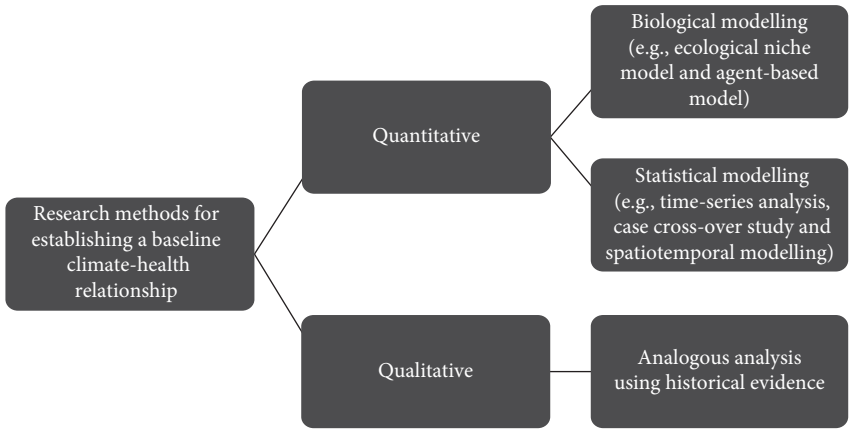


Fig. 15.5 The main research methods for establishing a baseline climate–health relationship

climate change in the past, present, and possible future. Relevant quantitative research includes two major types of methods: biological modelling and statistical (or empirical) modelling. The former includes process- or agent-based models such as the ecological niche model (Daszak et al. 2013) and the agent-based model (Bomblies and Eltahir 2009), while the latter includes time-series analysis (Yu et al. 2011; Huang et al. 2012a), case cross-over studies (Gomez-Acebo et al. 2010; Guo et al. 2011), spatio-temporal modelling (Kim et al. 2009), and survival analysis (Strand et al. 2012).

An estimation of the current and/or future burden of disease attributable to climate change should include five steps: (1) selection of risk-outcome pairs (e.g. the heatwave–death relation); (2) estimation of distributions of exposure to the risk factor (e.g. heatwave) in the population; (3) estimation of effect sizes, often relative risk per unit of exposure for each risk–outcome pair; (4) choice of an alternative (counterfactual) exposure distribution to which the current exposure distribution is compared (an optimum exposure distribution, termed the theoretical-minimum-risk exposure distribution, is often selected for this purpose); and (5) computation of burden attributable to each risk factor, including uncertainty from all sources (McMichael et al. 2004; Lim et al. 2012).

Epidemiologic studies are often required to quantify the fraction of current disease burden attributable to climate change and to project future climate change-related health effects based on the well-established climate–health relationship. However, estimating the full range of health impacts of climate change presents many challenges to conventional epidemiologic approaches (Campbell-Lendrum and Woodruff 2006; Patz et al. 2008). These include the

long timescale over which climate change will occur; long-term data on health outcomes (rarely available); the unprecedented large geographic areas potentially affected by climate change; the diversity of potential impacts on health; the complex interactions among demographic changes, socio-economic development, technological innovation, and other environmental drivers; and uncertainties in the role of adaptation and projections of future climate and socio-demographic changes. A recent scoping review indicates that, although climate change and health is a rapidly growing area of research, quantitative studies remain rare. Among recently published studies, there are gaps in adaptation research, and a deficit of studies in most developing regions (Hosking and Campbell-Lendrum 2012).

Modelling future health risks

Modelling future health risks requires the exposure measurement to be linked to a quantitative climate–health relationship (e.g. the change in disease rates per unit change in the climatic variable), for example, the increase in diarrhoea incidence in a population per year for each degree centigrade increase in average ambient temperature. This enables the calculation of a relative risk (i.e. proportional change) for the health outcomes under each of the various future climate change scenarios. The disease burden attributable to climate change is then estimated by multiplying this relative risk by the total burden of disease that would have been expected to occur in the absence of climate change (Campbell-Lendrum and Woodruff 2006).

To make inferences about current and future disease burdens, it is also necessary to account for the current and future influences of non-climatic factors such as socio-economic development. Non-climatic effects can be partly addressed by calculating relative risk estimates separately for populations with clearly different baseline disease burdens and vulnerabilities. Where possible, future relative risks should be applied to projections of disease burden that also account for changes in non-climatic influences over time, such as expected decreases in diarrhoea rates due to an improvement in water and sanitation services. Finally, changing socio-economic conditions and physiologic and behavioural adaptations will also affect the vulnerability of populations to the effects of climate change, that is, the relative risk as well as the baseline rate (Patz et al. 2008).

An estimation of the future burden of disease and premature death attributable to climate change for a specified year or time period in the future requires (1) a modelled scenario of global climate change; (2) estimations of population size and age structure by region, (3) estimations of future baseline (counterfactual) rates of disease incidence or premature death by region; and (4) assumptions

about the applicable relative risk (i.e. does it stay constant, increase, or decrease over time—in response to changes in that target population?). However, since global climate change is an 'exposure' that will extend and evolve over many decades, epidemiological studies must extend beyond the immediate future and also engage in attempts to estimate likely levels and patterns of health risks in future in response to plausible scenarios of climate change and its concomitants. We believe that the following issues need further discussion:

1. Models cannot predict with certainty.
2. Models necessarily entail simplified representation of a more complex, dynamic, relationship.
3. Models, while incomplete and simplified, are useful for (a) providing insights into mediating processes, (b) reducing complexities and background noise to a simpler mathematical representation of main elements, (c) providing indicative estimates of future impacts, and (d) enhancing communication to peers, the public, and policymakers.
4. The purpose of modelling and estimating future health risks in relation to plausible future climate scenarios is not to generate testable hypotheses but to conduct formal risk assessment that can then guide, and accelerate, the formulation of public policy.

An example: projecting future heat-related mortality

Heat-related mortality is a matter of great public health concern, especially in the light of climate change (Luber and McGeehin 2008). The IPCC indicates that hot weather is likely to increase future heat-related mortality. Projecting heat-related mortality under climate change scenarios will help decision-makers in planning adaptation strategies and communicating the future health risks of climate change to the public. However, understanding and managing uncertainty and complexity is the greatest challenge for projecting future heat-related mortality (Huang et al. 2011).

Essentially, projecting heat-related mortality under climate change models and scenarios requires an understanding of the baseline temperature–mortality relationships, and consideration of the future changes in climate, population, and acclimatization.

Baseline temperature–mortality relationships

Mortality projections are based on the exposure–response functions of temperature and mortality that are applied to climate change models and emission scenarios to estimate future heat-related mortality. Maximum temperature and mean temperature are commonly used measures of heat exposure. For

example, Dessai (2002, 2003) modelled the relationship between maximum temperature and excess deaths in Lisbon during the summer months in 1980–98 and then applied different climate and population change scenarios to the model to assess potential impacts on mortality in the 2020s and 2050s. Knowlton et al. (2007) projected the impacts of climate change on summer mortality using modelled daily mean temperatures for New York City. Others have used composite indices which examine the combined effects of ambient temperature, humidity, and other meteorological variables. For example, Jackson et al. (2010) examined the historical relationship between age-specific/cause-specific mortality rates and heat events at the 99th percentile of humidex values in Washington State.

Choosing a baseline time period for the temperature–mortality relationship is also important. Temperature–mortality relationships from the 1960s can be very different from those from the 2000s, even in the same city (Davis et al. 2003). Differences could be due to socio-economic development, demographic change, and/or population acclimatization. Differences in the time periods used to estimate the historical temperature–mortality relationships also make it difficult to compare projections between studies (Huang et al. 2011).

Climate change projections

Another fundamental issue for projecting heat-related mortality is the modelling of future climate. In the *Special Report on Emissions Scenarios* (SRES; Nakicenovic and Swart 2000), the IPCC defined a set of forty scenarios (termed SRES scenarios) that covered a wide range of the main driving forces of future greenhouse gas emissions. These scenarios are structured in four major families labelled A1, A2, B1, and B2. A1 represents rapid economic growth, global population peaking in mid-century, and rapid introduction of new and efficient technologies. A1 has three subgroups: A1FI (fossil intensive), A1T (non-fossil), and A1B (balanced). A2 represents high population growth, slow economic development, and slow technological change. B1 represents the same population growth as A1 but rapid changes in economic structures toward a service and information economy. B2 represents intermediate population and economic growth with local solutions to economic, social, and environmental sustainability. These emissions scenarios can be used to project future climates based on various general circulation models (GCMs; IPCC 2007).

For instance, Knowlton et al. (2007) considered two of the emission scenarios, A2 and B2, which assume relatively high and low future emissions, respectively. The authors used both scenarios to model daily mean temperatures in the 1990s and 2050s. Hayhoe et al. (2004) projected future climates based on the higher A1FI and lower B1 emission scenarios. Jackson et al. (2010) selected

three climate change scenarios for high (A1B), low (B1), and moderate (A1B and B1 combined) summer warming.

Climate change scenarios will determine the size of the predicted future heat-related mortality. Therefore, it is important to consider different emission scenarios in the impact assessment, offering a range of possible future climates and health impacts. The uncertainty associated with future emissions has been incorporated in the UKCP09 UK climate projections by providing probabilistic projections which correspond to each of three different emission scenarios: 'High', 'Medium', and 'Low' (Murphy et al. 2009). These scenarios correspond to three of the commonly used emission scenarios in SRES: A1FI, A1B, and B1, respectively.

Demographic changes

Challenges also arise from the uncertainties of future demographic changes that will modify the future sensitivity of populations to heat stress. Growing numbers of older adults will increase the proportion of the population at risk (Luber and McGeehin 2008). As well as having a diminished physiological ability to cope with heat, the elderly are more likely to live alone, have reduced social contacts, and experience poor health (Hajat et al. 2010).

To project the effects of climate change independent of the effects of population trends, one approach is to assume that the population size and age structure will remain constant. For example, Knowlton et al. (2007) assumed that population totals for each of the thirty-one counties in New York City, based on data obtained from the US Census 2000 survey, were held constant throughout the modelling period. Baseline mortality rates for all age groups were also held constant.

If susceptible populations are considered, then future demographic trends should be addressed. Jackson et al. (2010) obtained the county population estimates by age group for the years 2005–30. The population was held constant for the 2025 projection, allowing differences in excess deaths between years to be interpreted as the component due to climate change. Dessai (2003) estimated population scenarios for Lisbon in line with the SRES. The population growth rates from each SRES storyline were applied to the 1990 Lisbon population to produce future population figures until 2100, and the median population from these calculations was used.

Population acclimatization

How populations may acclimatize to elevated temperatures over time is another issue affecting mortality projections. Acclimatization can be a physiological process of humans adjusting to changes in their environment. People may also

adapt to extreme heat through increased use of air conditioning, modified behaviour patterns, and improved building designs and urban planning (Huang et al. 2011).

One approach is to assume that no acclimatization takes place in the future. For example, Baccini et al. (2011) argued that epidemiological evidence of the extent to which short- or long-term acclimatization alters mortality risk is limited and sometimes discordant. Therefore, for their projections, no acclimatization was assumed and thus no future change in the temperature–mortality relationship.

To incorporate acclimatization, one approach is to use the exposure–response curves from analogue cities. These analogues represent cities whose present climate best approximates the estimated future climate of a target city. For example, Knowlton et al. (2007) modelled acclimatization in New York City using a temperature–mortality response function derived for Washington, DC, and Atlanta, GA, both of which had mean summer temperatures for 1973–94 that were within approximately 1°F of projected temperatures for the New York City region in the 2050s. Another approach involves the use of analogue summers from the same city to model population acclimatization. Hayhoe et al. (2004) used analogue summers in which future acclimatization was based on the temperature–mortality relationships only from the hottest summers on record. Cheng et al. (2009) identified the five hottest and five coolest summers during 1953–2000 and attributed the differences in daily mean deaths between the hottest and coolest summers to acclimatization. Others have accounted for acclimatization by shifting current temperature–mortality relationships to the future. Using this method, the heat threshold increases with time but the slope of the temperature–mortality relationship remains unchanged. Dessai (2003) assumed that complete acclimatization to an extra 1°C warming in maximum temperature is reached every three decades. Gosling et al. (2009) considered three possibilities of future acclimatization: no acclimatization, acclimatization to an increase of 2°C, and acclimatization to an increase of 4°C.

Efforts to better understand how climate change will affect population health, especially among the most vulnerable groups, are necessary. Given uncertainties in our understanding of the future population vulnerability to heat, it is important to use various methods to capture a plausible range of the health impacts of climate change. Although the methods used for projections are still in their early stages and have limitations, the need for evidence-based assessments of future health impacts of climate change is urgent. Such research will significantly contribute to assessing and managing the potential impacts of climate change on population health and well-being (Huang et al. 2013).

Climate change and health: adaptation strategies

The research-and-policy discussion of climate change and its impacts is paying increasing attention to the need for adaptive strategies to lessen risks. The term 'adaptive strategies' comes primarily from the well-recognized work of the IPCC and is the counterpoint to the 'mitigation strategies' that seek to reduce greenhouse gas emissions.

Following the notorious heat wave of August 2003 in Europe, there was an assessment of local and national actions introduced in France, including heat wave warning systems and surveillance of health events and environmental indices (Pascal et al. 2012).

The amount of attention given to adaptation has increased for two main reasons:

1. the recognition that climate change already exists and is affecting various physical, biological, social, and human health outcomes and that most of the impacts are adverse; and
2. the realization that climate variability will be an increasingly important aspect of the process of climate change and will present a range of often extreme stresses and threats to vulnerable systems.

Adaptation can occur via both *planned adaptation* (i.e. activities deliberately conducted by individuals, communities, or government) and *autonomous adaptation* (i.e. spontaneous responses to changes in the climate—physiological, behavioural, cultural, and/or institutional). Epidemiological assessments of the effectiveness of adaptive strategies must take into account the effect of coexistent environmental and socio-economic stresses that influence individual and community adaptive capacity.

In principle, there are three types of adaptive strategies:

- Strategies to reduce exposures to climate-related health hazards (e.g. exposures to extremes of heat, or to infectious agents).
- Strategies to increase the resilience (reduce the susceptibility) of the exposed group or community.
- Strategies to speed or enhance the recovery from adverse health impacts.

These are the three essential components of vulnerability: exposure, sensitivity, and the capacity to respond. Essentially, adaptation to the impacts of climate change means a reduction in vulnerability. This involves changes in attitudes and perception, social structure changes that mediate risk, and an improvement in health-care services and public health programmes (O'Brien et al. 2012).

It is relevant to note that most of the health impacts of climate change will result from the intensification of naturally occurring climatic exposures and the

associated changes in environmental, ecological, and social conditions. The existing health systems and public health practices should therefore provide a good starting point for dealing with many of the anticipated health impacts (Frumkin et al. 2008).

Many adaptive strategies will be amenable to formal epidemiological study for the evaluation of their effectiveness, cost-effectiveness, equity, time-course, and other policy-relevant characteristics. Study designs will encompass methods that are non-experimental, quasi-experimental (e.g. comparing heat wave mortality reductions achieved by differing heat-alert systems in adjoining states/jurisdictions), and controlled experimental. In addition to empirical studies of the outcome of adaptive strategies, there will also be a need for behavioural epidemiological studies of what best facilitates communities and individuals to participate in, or respond to, adaptive strategies, and social-epidemiological and policy analytic studies of social, cultural, and policy barriers to the implementation of adaptive strategies.

The example given in Fig. 15.6 illustrates a range of adaptive strategies that might be applied to lessen the health risks from exposures to extremes of heat.

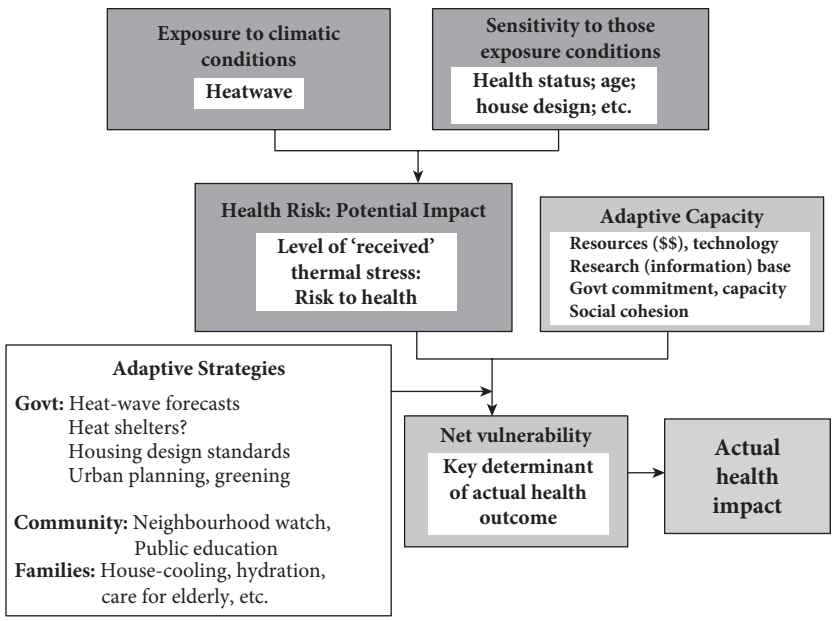


Fig. 15.6 Illustrative example of the components of 'vulnerability' as determinant of the actual health impact of a climate-related exposure.

Adapted from Vulnerability Model, Potsdam Institute for Climate Impact Research website, Copyright © 2014. Accessed at <https://www.pik-potsdam.de>.

The choice of adaptive strategies will often need to take account of cost–benefit considerations.

Adaptive strategies: differences in vulnerability

Vulnerability issues differ between regions and between developed and developing countries. Many developing countries exist at low latitudes and have relatively high average temperatures. This profile, which may include exposure to cyclones, increases the likelihood of droughts, floods, and outbreaks of infectious diseases. Populations in low-income, less urbanized countries rely more on climate-sensitive natural resources such as forests, agricultural lands, and fisheries for subsistence or employment. Poorer subgroups often live and work in environmentally high-risk areas, such as flood-prone river plains, unproductive agricultural lands, or steep slopes that are susceptible to mudslides. Additional stresses, such as HIV/AIDS, poverty, inequality, and political instability, exacerbate the above situations by amplifying the inherent susceptibility of individuals and communities—which heightens the adverse impacts of climate change and reduces the community's ability to adapt.

The IPCC (2007) *Fourth Assessment Report* emphasizes that the adverse health impacts of climate change will fall primarily on low-income, poorly resourced, and geographically vulnerable populations—especially impoverished slum-dwelling populations, subsistence farmers, the homeless, and resource-dependent communities (e.g. fishermen), in both developing and developed countries.

In Australia, for example, groups likely to be vulnerable to climate change include

- persons living in regions where climate-sensitive infectious diseases may tend to spread;
- rural (especially farming) communities in southern and eastern Australia (exposed to long-term drying conditions);
- older and frailer persons, especially in relation to heat waves;
- coastal-resident communities;
- remote indigenous communities facing heat, drying, water shortages, and the loss of traditional food species; and
- those living in current and potential cyclone-risk zones.

In practice, the actual choice and form of adaptive interventions depends on the scale under consideration (i.e. global, regional, or local), the type of health risk, the time-frame, and the resources available. This has implications for the choice

of research methods by which such intervention strategies are evaluated (Menne and Ebi 2006).

Early warning systems and vulnerability mapping are well-recognized measures of adaptation for the protection of health. The latter can be made with a specific focus (e.g. disease specific; Dickin et al. 2013) or with a broader perspective (Thornton et al. 2008; Confalonieri et al. 2009, 2014).

Basic public health functions in relation to climate change: risk assessment, surveillance, and information dissemination

A range of public health functions is required in response to climate change (Campbell-Lendrum et al. 2007; Frumkin et al. 2008). Many will require inputs from the research discipline of epidemiology: concepts, measures, and methods of evaluation. In particular, there will be a need to

1. document and communicate, to public and policymakers, the actuality of health risks due to climate change—national and (appropriate) sub-national formal health-risk assessments should be carried out, to identify the main health risks, their likely chronology, and the vulnerable subpopulations;
2. anticipate where health impacts are most likely to appear and ensure there is good, continuing, health outcome surveillance directed at those 'pressure points';
3. develop methods of estimating causal attribution such that public and policymakers can be advised of the likely contribution of underlying climate change to the impacts of otherwise 'natural' events (e.g. Hurricane Katrina and Europe's 2003 heat wave)—this also requires appropriate handling and communication of the complex issue of uncertainty;
4. enhance the overall repertoire of prevention (adaptive) strategies (e.g. early warning systems for health-threatening events, community alerts for fragile older persons, better surveillance systems for detecting shifts in infectious disease patterns, enhanced disaster response preparedness, and food supplementation systems)—in addition, it will be necessary to accrue experience and knowledge about the range of options for these and other risk-lessening adaptive interventions and to evaluate the options for specified populations/communities in terms of averted disease/death/disability burden and cost-benefit profile; and
5. undertake systematic updating of scenario-based health-risk assessments in relation to plausible climate change futures—this will assist and update the understanding, by both general public and policymakers, of the likely future health risks resulting from plausible future social and economic changes.

The US Centers for Disease Control (CDC) website *Climate and Health* (<http://www.cdc.gov/climateandhealth/>; updated in 2014) has a good discussion about the public health functions required for responding to climate change, as well as a number of useful links.

Assessment students' achievements

These are early days in the teaching of this new and evolving topic. Obvious options for the assessment of student performance include

- a critical review of a nominated published paper that exemplifies one or more of the research categories described in 'Discussing the research tasks for epidemiology';
- an essay that provides a review and critical commentary about the types of research methods needed in this research domain and what this means for the future development of epidemiological research; and
- a draft research proposal to carry out a study under one of the four core categories discussed in 'Health impacts of climate change: what types of epidemiological study?'. This could be done by a single student or by a small group.

Conclusion

Traditional quantitative epidemiological methods play a role in the underlining of climate–health linkages but more robust analytical approaches are still needed to clarify the complexities of many process involved in the impacts of climate factors upon human population health. Human health is—and will be—affected by climate via pathways involving large-scale and long-term exposures, with multiple effects on natural and human systems.

The immediate public health impacts of extreme climate events are readily imaginable or apparent. In the long term, a greater health hazard for many populations will arise from disturbances to environmental, ecological, and social living conditions. This mix of direct and indirect risks to health, and immediate and eventual health risks, poses a range of research challenges to epidemiologists. A major challenge ahead is to fill important gaps in the knowledge of the diverse and complex pathways through which human health is affected by the changing global climate system. This is especially true in developing countries, where the most vulnerable people live and which often have limited scientific research capacity.

The case for adaptive strategies to lessen health risks from climate change is clear. This, too, poses a responsibility and challenge for epidemiologists, for the

public health system, and for the health sector at large in its dealings with other sectors and social agencies.

Overall, there is a great need for epidemiologists to learn about and engage in addressing this momentous issue. There is important research to be done—much of it at unusual scale and with unfamiliar collaborating disciplines. All such research will contribute to enriching and informing the formulation of public policy on one of the great issues of the age.

References

Alonso, D., Bouma, M. J., and Pascual, M. (2011) Epidemic malaria and warmer temperatures in recent decades in an east African highland. *Proceedings of the Royal Society B: Biological Sciences*, **278**: 1661–9.

Anenberg, S. C. et al. (2012). Global air quality and health co-benefits of mitigating near-term climate change through methane and black carbon emission controls. *Environmental Health Perspectives*, **120**: 831–9.

Baccini, M., Kosatsky, T., Analitis, A., Anderson, H. R., D'Ovidio, M., Menne, B., Michelozzi, P., Biggeri, A., and the PHEWE Collaborative Group. (2011) Impact of heat on mortality in 15 European cities: attributable deaths under different weather scenarios. *Journal of Epidemiology and Community Health*, **65**: 64–70.

Ballard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., and Courchamp, C. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**: 356–77.

Bandyopadhyay, S., Kanji, S., and Wang, L. (2012) The impact of rainfall and temperature variation on diarrheal prevalence in Sub-Saharan Africa. *Applied Geography*, **33**: 63–72.

Béguin, A., Hales, S., Rocklöv, J., Åström, C., Louis, V. R., and Sauerborn, R. (2011) The opposing effects of climate change and socio-economic development on the global distribution of malaria. *Global Environmental Change*, **21**: 1209–14.

Berry, H. L., Bowen, K., and Kjellstrom, T. (2010) Climate change and mental health: a causal pathways framework. *International Journal of Public Health*, **55**: 123–132.

Bomblies, A and Eltahir, E. B. (2009) Assessment of the impact of climate shifts on malaria transmission in the Sahel. *EcoHealth*, 6: 426–37.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984) *Classification and Regression Trees*. Belmon, CA: Wardworth.

Campbell-Lendrum, D., Corvalan, C., and Neira, M. (2007) Global climate change: implications for international public health policy. *Bulletin of the World Health Organization*, **85**: 235–7.

Campbell-Lendrum, D and Woodruff, R. (2006) Comparative risk assessment of the burden of disease from climate change. *Environmental Health Perspectives*, **114**: 1935–41.

Chaves, L. F. and Koenraadt, C. J. (2010) Climate change and highland malaria: fresh air for a hot debate. *Quarterly Review of Biology*, **85**: 27–55.

Chaves, L. F. and Pascual, M. (2006) Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. *PLoS Medicine*, **3**: e295.

Cheng, C. S. et al. (2009) Differential and combined impacts of extreme temperatures and air pollution on human mortality in south-central Canada. Part II: future estimates. *Air Quality, Atmosphere and Health*, **1**: 223–35.

Chinga-Alayo, E., Huarcaya, E., Nasarre, C., Del Aguila, R., and Llanos-Cuentas, A. (2004) The influence of climate factors on the epidemiology of bartonellosis in Ancash, Peru. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **298**: 116–24.

Confalonieri, U. E. C., Lima, A. C. L., Brito, I., and Quintão, A. F. Q. (2014) Social, environmental and health vulnerability to climate change in the Brazilian Northeastern Region. *Climatic Change*, **127**: 123–37.

Confalonieri, U. E. C., Marinho, D. P., and Rodriguez, R. E. (2009) Public health vulnerability to climate change in Brazil. *Climate Research*, **40**: 175–86.

Confalonieri,U. Menne, B., Akhtar, R., Ebi, K. L., Hauengue, M., Kovats, R. S., Revich, B., and Woodward, A. (2007) 'Human health', in M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, eds, *Climate Change 2007: Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, pp. 391–431.

Craig, M. H., Snow, R. W., and le Sueur, D. (1999) A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, **15**: 105–11.

Daszak, P., Cunningham, A., and Hyatt, A. D. (2000) Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science*, **287**: 443–9.

Daszak, P., Zambrana-Torrelio, C., Bogich, T. L., Fernandez, M., Epstein, J. H., Murray K. A., and Hamilton, H. (2013) Interdisciplinary approaches to understanding disease emergence: the past, present, and future drivers of Nipah virus emergence. *Proceedings of the National Academy of Sciences USA*, **110** Suppl. 1: 3681–8.

Davis, R. E., Knappenberger, P. C., Michaels, P. J., and Novicoff, W. M. (2003). Changing heat-related mortality in the United States. *Environmental Health Perspectives*, **111**:1712.

Dessai S. (2002). Heat stress and mortality in Lisbon Part I. Model construction and validation. *International Journal of Biometeorology*, **47**: 6–12.

Dessai S. (2003). Heat stress and mortality in Lisbon Part II. An assessment of the potential impacts of climate change. *International Journal of Biometeorology*, **48**: 37–44.

Diamond, J. (2005) Collapse: *How Societies Choose to Fail or Survive*. London: Penguin, Allen Lane.

Dickin, S. K., Schuster-Wallace, C. J., and Elioh, S. (2013) Developing a vulnerability mapping methodology: applying the water-associated disease index to dengue in Malaysia. *PLoS ONE*, **8**: e63584.

Dodman, D., Mitlin, D., and Co, J. R. (2010) Victims to victors, disasters to opportunities: community-driven responses to climate change in the Philippines. *International Development Planning Review*, **32**: 1–26.

Earnest, A., Tan, S. B., and Wilder-Smith, A. (2012) Meteorological factors and El Niño-Southern Oscillation are independently associated with dengue infections. *Epidemiology and Infection*, **140**: 1244–51.

Ebi, K. L., Hartman, J., Chan, N., McConnell, J., Schlesinger, M., and Weyant, J. (2005) Climate suitability for stable malaria transmission in Zimbabwe under different climate change scenarios. *Climatic Change*, **73**: 375–93.

Ebi, K. L. and Schmier, J. K. (2005) A stitch in time: improving public health early warning systems for extreme weather events. *Epidemiologic Reviews*, **27**: 115–21.

Field, C. B. et al., eds. (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. New York: Cambridge University Press.

Forget, G., and Lebel, J. (2001) An ecosystem approach to human health. *International Journal of Occupational and Environmental Health*, **7**: S1–38.

Friel, S., Dangour, A. D., Garnett, T., Lock, K., Chalabi, Z., Roberts, I., Butler, A., Butler, C. D., Waage, J., McMichael, A. J., and Haines, A. (2009) Public health benefits of strategies to reduce greenhouse-gas emissions: food and agriculture. *Lancet*, **374**: 2016–25.

Frumkin, H., McMichael, A. J., and Hess, J. J., eds. (2008) Climate change and the health of the public. *American Journal of Preventive Medicine*, **35**: 401–538.

Gomez-Acebo, I., Dierssen-Sotos, T., and Llorca, J. (2010) Effect of cold temperatures on mortality in Cantabria (Northern Spain): a case-crossover study. *Public Health*, 124: 398–403.

González, C., Wang, O., Strutz, S. E., González-Salazar, C., Sánchez-Cordero, V., and Sarkar, S. (2010) Climate change and risk of leishmaniasis in North America: predictions form ecological niche models of vector and reservoir species. *PLoS Neglected Tropical Diseases*, **4**: e585.

Gosling, S., McGregor, G., and Lowe, J. (2009) Climate change and heat-related mortality in six cities Part 2: climate model evaluation and projected impacts from changes in the mean and variability of temperature with climate change. *International Journal of Biometeorology*, **53**: 31–51.

Guo, Y., Barnett, A., Yu, W., Pan, X., and Tong, S. (2011) The impact of temperature on mortality in Tianjin, China: a case-crossover design with a distributed lag non-linear model. *Environmental Health Perspectives*, **119**: 1719–25.

Haines, A., and Cassel, A. (2004) Can the millennium development goals be attained? *British Medical Journal*, **329**: 394–7.

Haines, A. et al. (2009) Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers. *Lancet*, **374**: 2104–14.

Haines, A., Kovats, R., Campbell-Lendrum, D., and Corvalan, C. (2006) Climate change and human health: impacts, vulnerability, and mitigation. *Lancet*, **367**: 2101–9.

Hajat, S., O'Connor, M., and Kosatsky, T. (2010) Health effects of hot weather: from awareness of risk factors to effective health protection. *Lancet*, **375**: 856–63.

Hastie, T. and Tibshirani, R. (1990) *Generalized Additive Models*. London: Chapman & Hall/CRC.

Hayhoe, K. et al. (2004). Emissions pathways, climate change, and impacts on California. *Proceedings of the National Academy of Sciences USA*, **101**: 12422–7.

Holt, A. C., Salkeld, D. J., Fritz, C. L., Tucker, J. R., and Gong, P. (2009) Spatial analysis of plague in California: niche modeling predictions of the current distributions and potential response to climate change. *International Journal of Health Geographics*, **8**: 35–51.

Hosking, J., and Campbell-Lendrum, D. (2012) How well does climate change and human health research match the demands of policymakers? A scoping review. *Environmental Health Perspectives*, 120: 1076–82.

Huang, C., Barnett, A. G., Wang, X., and Tong, S. (2012a) Effects of extreme temperatures on years of life lost for cardiovascular deaths: a time series study in Brisbane, Australia. *Circulation: Cardiovascular Quality and Outcomes*, **5**: 609–14.

Huang, C., Barnett, A. G., Wang, X., and Tong, S. (2012b) The impact of temperature on years of life lost in Brisbane, Australia. Nature Climate Change, 2: 265–70.

Huang, C., Barnett, A. G., Wang, X., Vaneckova, P., FitzGerald, G., and Tong, S. (2011) Projecting future heat-related mortality under climate change scenarios: a systematic review. *Environmental Health Perspectives*, **119**: 1681–90.

Huang, C., Barnett, A., Xu, Z., Chu, C., Wang, X., Turner, L., and Tong, S. (2013) Managing the health effects of temperature in response to climate change: challenges ahead. *Environmental Health Perspectives*, **121**: 415–19.

Intergovernmental Panel on Climate Change (IPCC). (2007) *Fourth Assessment Report (AR4).* <http://ipcc.ch/report/ar4/>, accessed 7 November 2014.

Intergovernmental Panel on Climate Change (**IPCC**). (2013) *Climate Change 2013: The Physical Science Basis Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Jackson, J. E., Yost, M. G., Karr, C., Fitzpatrick, C., Lamb, B. K., Chung, S. H., Chen, J., Avise, J., Rosenblatt, R. A., and Fenske, R. A. (2010) Public health impacts of climate change in Washington State: projected mortality risks due to heat events and air pollution. *Climatic Change*, **102**: 159–86.

Keim, M. E. (2008) Building human resilience: the role of public health preparedness and response as an adaptation to climate change. *American Journal of Preventive Medicine*, **35**: 508–16.

Kim, J., Lawson, A. B., McDermott, S., Aelion, C. M. (2009) Variable selection for spatial random field predictors under a Bayesian mixed hierarchical spatial model. *Spatial and Spatio-Temporal Epidemiology*, **1**: 95–102.

Knowlton, K., Lynn, B., Goldberg, R. A., Rosenzweig, C., Hogrefe, C., Rosenthal, J. K., and Kinney, P. L. (2007) Projecting heat-related mortality impacts under a changing climate in the New York City region. *American Journal of Public Health*, **97**: 2028–34.

Krieger, N. (1994) Epidemiology and the web of causation: has anyone seen the spider? *Social Science and Medicine*, **39**: 887–903.

Krieger, N. (2001a) A glossary for social epidemiology. *Journal of Epidemiology and Community Health*, **55**: 693–700.

Krieger, N. (2001b) Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology*, **30**: 668–77.

Lim, S. S. et al. (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**: 2224–60.

Lindgren, E. and Gustavson, R. (2001) Tick-borne encephalitis in Sweden and climate change. *Lancet*, **358**: 16–18.

Lindgren, E. and Naucke, T. (2006) 'Leishmaniasis: influences of climate and climate change epidemiology, ecology and adaptation measures', in **B. Menne** and **K. Ebi**, eds, *Climate Change and Adaptation Strategies for Human Health*. Darmstadt: Steinkopff, pp. 131–56.

Lindgren, E., Talleklint, L., and Polfeldt, T. (2000) Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick *Ixodes ricinus*. *Environmental Health Perspectives*, **108**: 119–23.

Luber, G. and McGeehin, M. (2008) Climate change and extreme heat events. *American Journal of Preventative Medicine*, **35**: 429–35.

Mangal, T.D., Paterson, S., and Fenton, A. (2008) Predicting the impact of long-term temperature changes on the epidemiology and control of schistosomiasis: a mechanistic model. *PLoS ONE*, **3**: e1438.

McCullagh, P. and Nelder, J. (1989) *Generalised Linear Models* (2nd edn). London: Chapman & Hall/CRC.

McLaughlin, J. B., DePaola, A., Bopp, C. A., Martinek, K. A., Napolilli, N. P., Allison, C. G., Murray, S. L., Thompson, E. C., Bird, M. M., and Middaugh, J. P. (2005) Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. *New England Journal of Medicine*, **353**: 1463–70.

McMichael, A. J. (1999) Prisoners of the proximate: loosening the constraints one epidemiology in an era of change. *American Journal of Epidemiology*, **149**: 887–97.

McMichael, A. J. (2001) *Human Frontiers, Environments and Disease: Past Patterns, Future Uncertainties*. Cambridge: Cambridge University Press.

McMichael, A. J. (2012) Insights from past millennia into climatic impacts on human health and survival. *Proceedings of the National Academy of Sciences USA*, **109**: 4730–7.

McMichael, A. J. (2013) Globalization, climate change, and human health. *New England Journal of Medicine*, 368: 1335–43.

McMichael, A. J. et al. (2004) 'Global climate change', in M. Ezzati, A. D. Lopez, A. Rodgers, and C. J. L. Murray, eds, *Comparative Quantification of Health Risks: Global and Regional Burden of Disease due to Selected Major Risk Factors*. Geneva: WHO, pp. 1543–1649.

McMichael, A. J., Woodruff, R., and Hales, S. (2006) Climate change and human health: present and future risks. *Lancet*, 367: 859–69.

Menne, B. and Ebi, K. L. (2006) *Climate Change and Adaptation Strategies for Human Health in Europe*. Darmstadt: Steinkopff-Verlag.

Millennium Ecosystem Assessment. (2005) *Ecosystems and Human Well-Being: Current State and Trends, Volume 1.* <http://www.millenniumassessment.org/en/Condition.html>, accessed 8 November 2014.

Murphy, J. M. et al. (2009) *UK Climate Projections Science Report: Climate Change Projections*. Exeter: Meteorological Office Hadley Centre.

Nakicenovic, N. and Swart, R., eds. (2000). *Special Report on Emissions Scenarios*, Cambridge: Cambridge University Press.

O'Brien, K., Pelling, M., Patwardhan, A., Hallegatte, S., Maskrey, A., Oki, T., Oswald-Spring, U., Wilbanks, T., and Yanda, P. Z. (2012) 'Toward a sustainable and resilient future', in: Field, C. B. et al., eds, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. New York: Cambridge University Press, pp. 437–86.

Parmesan, C. and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**: 37–42.

Pascal, M., Laaidi, K., Wagner, V., Ung, A. B., Smaili, S., Fouillet, A., Caserio-Schönemann, C., and Beaudeau, P. (2012) How to use near real-time health indicators to support decision-making during a heat wave: the example of the French heat wave warning system. *PLoS Currents*, **4**:e4f83ebf72317d.

Pascual, M., Ahumada, J. A., Chaves, L. F., Rodó, X., and Bouma, M. (2006) Malaria resurgence in the East African highlands: temperature trends revisited. *Proceedings of the National Academy of Sciences USA*, **103**: 5829–34.

Patz, J., Campbell-Lendrum, D., Gibbs, H., and Woodruff, R. (2008) Health impact assessment of global climate change: expanding on comparative risk assessment approaches for policy making. *Annual Review of Public Health*, **29**: 27–39.

Patz, J., Campbell-Lendrum, D., Holloway, T., and Foley, J. A. (2005) Impact of regional climate change on human health. *Nature*, **438**: 310–17.

Pearce, N. (2000) The ecological fallacy strikes back. *Journal of Epidemiology and Community Health*, **54**: 326–7.

Peterson, A.T. (2009) Shifting suitability for malaria vectors across Africa with warming climates. *BMC Infectious Diseases*, **9**: 59–64.

Rose, G. (1985) Sick individuals or sick populations? *International Journal of Epidemiology*, **14**: 32–8.

Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, W. D., Howe, H. L., and Wilhelm, J.L. (1996) Heat-related deaths during the July 1995 heat wave in Chicago. *New England Journal of Medicine*, **335**: 84–90.

Stanke, C., Kerac, M., Proudhomme, C., Medlock, J., and Murray, V. (2013) Health effects of drought: a systematic review of the evidence. *PLoS Currents*, **5**: ecurrents.dis.7a2cee9e980f91ad7697b570bcc4b004.

Stern, N. (2007) *The Economics of Climate Change (The Stern Review)*. Cambridge: Cambridge University Press.

Strand, L., Barnett, A., and Tong, S. (2012) Maternal exposure to ambient temperature and the risk of preterm birth and stillbirth in Brisbane, Australia. *American Journal of Epidemiology*, **175**: 99–107.

Susser, M. and Susser, E. (1996) Choosing a future for epidemiology: II. From black boxes to Chinese boxes and eco-epidemiology. *American Journal of Public Health*, **86**: 674–7.

Tan, J., Zheng, Y., Song, G., Kalkstein, L. S., Kalkstein, A. J., and Tang, X. (2007) Heat wave impacts on mortality in Shanghai, 1998 and 2003. *International Journal of Biometeorology*, **51**: 193–200.

Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A. P., and Palmer, T. N. (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**: 576–9.

Thornton, P. et al. (2008). Climate change and poverty in Africa: mapping hotspots of vulnerability. *African Journal of Agriculture and Resource Economics*, **2**: 24–44.

UNEP. (2007) *Sudan: Post-Conflict Environmental Assessment*. <http://www.unep.org/sudan/post-conflict>, accessed 9 November 2014.

Van Lieshout, M. Kovats, R. S., Livermore, M. T. J., and Martens, P. (2004) Climate change and malaria: analysis of the SRES climate and socio-economic scenarios. *Global Environmental Change*, **14**: 87–99.

Vandentorren, S., Suzan, F., Medina, S., Pascal, M., Maulpoix, A., Cohen, J. C., and Ledrans, M. (2004) Mortality in 13 French cities during the August 2003 heat wave. *American Journal of Public Health*, **94**:1518–20.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., Fromentin, J. M., Hoegh-Guldberg, O., and Bairlein, F. (2002) Ecological responses to recent climate change. *Nature*, **416**: 389–95.

Wu, P., Guo, H. R., Lung, S. C., Lin, C. Y., and Su, H. J. (2007) Weather as an effective predictor of the occurrence of dengue fever in Taiwan. *Acta Tropica*, **103**: 50–7.

WWF. (2006). *Living Planet Report 2006.* <http://assets.panda.org/downloads/living_planet_report.pdf>, accessed 9 November 2014.

Yang, G. J., Vounatsou, P., Zhou, X. N., Tanner, M., and Utzinger, J. (2005) A potential impact of climate change and water resource development on the transmission of *Schistosoma japonicum* in China. *Parasitologica*, **47**: 127–34.

Yu, W. W., Hu, W., Guo, Y. M., Mengersen, K., Pan, X. C., Connell, D., and Tong, S. (2011) The time course of temperature effects on cardiovascular mortality in Brisbane, Australia. *Heart*, **97**:1089–93.

Zhou, X.-N., Yang, G. J., Yang, K., Wang, X. H., Hong, Q. B., Sun, L. P., Malone, J. B., Kristensen, T. K., Bergquist, N. R., and Utzinger, J. (2008) Potential impact of climate change on schistosomiasis transmission in China. *American Journal of Tropical Medicine and Hygiene*, **78**: 188–94.

Part 3

Outcome-oriented epidemiology

Chapter 16

Infectious disease epidemiology

Marc Lipsitch

Introduction to infectious disease epidemiology

Infectious disease epidemiology courses in practice concentrate almost exclusively on communicable diseases—those transmitted from person to person. Such courses are generally taught in one of two ways. The first approach is to define the biological basis of infectious diseases and then follow with sessions on the major types of infectious diseases as classified by route of transmission (airborne, close contact, vector, sexual), the major pathologic consequence or system affected (respiratory, gastrointestinal, genitourinary, oncogenic infections), and/or by the type of infectious agent involved (e.g. helminths, fungi, protozoa, bacteria, viruses, or prions). Additional sessions might focus on technical and practical topics such as outbreak investigation, infection control, study design, and the molecular epidemiology of pathogens. In such courses natural history, descriptive epidemiology, and the biology of host and pathogen take centre stage, and the differences among infections are emphasized. Several major textbooks of infectious disease epidemiology follow this approach (Thomas and Weber 2001; Nelson and Masters Williams 2006). A second approach is to focus on the common factors that unite infectious (more specifically, communicable) diseases, using particular diseases as examples to illustrate the principles of communicable disease epidemiology while abandoning any effort to be comprehensive in covering the major types of infectious diseases.

Both approaches are useful, particularly for students without medical training (which often provides a tolerable, if not ideal, introduction to the topics covered in the first approach) and, ideally, a student with a deep interest in the subject will be exposed to each of them. However, this chapter describes how a course might be taught following the second approach, which emphasizes principles and commonalities at the expense of breadth of coverage. The goal is to expose students to analytic approaches and ways of thinking about infectious diseases that will serve them well in considering any communicable disease or, indeed, in assessing the possible infectious etiology of diseases whose

causes are unknown (Knobler et al. 2004). An excellent recent textbook by Vynnycky and White (2010) has been a useful adjunct to teaching a course along the lines described here; notwithstanding the phrase 'mathematical modelling' in the title, it is a rather biologically and epidemiologically motivated treatment of the topic. Supplementation from the text by Giesecke (2001), from the texts mentioned above, and with papers from the primary literature may be helpful. The modern 'bible' of infectious disease dynamics is the book by Anderson and May (1991), which may be difficult reading for non-mathematical students but is an invaluable resource for teaching and research.

Teaching objectives

Following the course described here, a student should be able to

- correctly use terminology associated with infectious diseases and their transmission;
- assess the evidence in favour of an infectious cause for a particular disease;
- describe secular trends in infectious disease morbidity and mortality;
- define, use, and make simple calculations concerning key concepts in infectious disease epidemiology, including the basic reproductive number, herd immunity, epidemic curves, and recurrent epidemics in open populations;
- understand the qualitative impact of heterogeneity in disease susceptibility and infectiousness risk on the probability of epidemics, the age structure of disease incidence, and the concentration of disease risk;
- describe the individual-level and population-level impacts of disease-control measures such as vaccination, antimicrobial treatment, and hygiene (including condom use);
- list the key determinants of the rate/extent of spread of a communicable disease;
- describe the impact of these key determinants; and
- use simple, compartmental models of infectious disease transmission to assess the impact of variations in disease parameters on the transmission of the disease.

Key themes that should be emphasized throughout and revisited through case studies include

- the interacting roles of biology (host and pathogen characteristics, and evolution) and behaviour in determining the transmission of infectious diseases;

- the distinctive effects of medical or public health interventions to control infectious diseases on the individual 'recipient' of the intervention and the community as a whole—often extending benefits (reduced transmission) to those who do not receive the intervention but in a few instances placing them at increased risk; and
- the use of mathematical models as tools to understand infectious disease dynamics.

Teaching method

This course may be taught as a series of lectures of one to two hours each; approximate lengths are suggested for each lecture but may be modified according to the interest and expertise of the lecturer and the level of the students.

A central theme in the course is that every characteristic or action of one individual in the population that affects his ability to become infected with and/or transmit a communicable disease has consequences for others in the population. This is the key distinguishing feature of communicable diseases and affects nearly every aspect of their epidemiology. After this course, students should almost instinctively ask themselves when considering a public health intervention for a communicable disease what its effects would be both on the targeted individuals and on the population as a whole, including those not targeted.

In technical subjects, it is surprisingly easy for students to acquire the ability to manipulate equations and answer quantitatively challenging exam questions without acquiring a qualitative or intuitive understanding of the subject (Mazur 2005). In lecturing, in writing exam questions, and in preparing guiding questions to help students read the primary literature, it is valuable to pose simple questions that have precise answers but require qualitative understanding. Often these may be in the form of applied questions, such as 'If you were weighing a decision to vaccinate your child and you believed that vaccine carried a certain risk of an adverse event, which numbers from the following table would you use and how would you use them?' (I have previously used this question in relation to Table 2 in Feikin et al. 2000 during a group discussion of this paper). Sometimes the questions may be more abstract: when studying age-seroprevalence data, ask students to sketch the curve one would expect for a sexually transmitted disease that has been in the population for a long time, then one for a childhood disease that was eradicated thirty years ago, etc. In some cases these qualitative questions are considerably harder than questions involving more complex calculations, just as Mazur (2005) found for his physics students, in that they motivate and test a different form of understanding.

Teaching content

What follows is an outline of a course that could be taught to students with a limited background in epidemiology; a target audience might be physicians or recent college graduates with some biological background.

The level of the course should be adapted to student abilities and interests. Students with excellent quantitative skills or a background in another population science (demography, ecology) may be able to move rapidly through the conceptual aspects of the early lectures. Three options are possible in such circumstances: (1) compress the presentation of this material; (2) add a focus on data sources, data analysis, and the biological basis of the phenomena, for example, immunity; and (3) add quantitative detail, for example, at the level of Anderson and May 1991 or Keeling and Rohani 2007; or, for mathematicians, Diekmann and Heesterbeek 2000.

Introduction, terminology, trends, techniques, and causality (two to three hours)

To capture student interest before launching into the dull preliminaries—while giving a sense of the approach of the course—it is suggested to open the class with a series of 'big questions' about infectious diseases and their common features. The idea is that students should have some insight into (or in some cases, the answer to) these questions by the end of the class. Common features of communicable diseases include seasonal, recurrent epidemics (sometimes annual, sometimes every several years; Anderson and May 1991); age-distribution of attack rates (Anderson and May 1991); and rates of change of drug resistance in hospital- and community-acquired pathogens (Lipsitch 2001b). Students may be presented with these phenomena and asked to speculate about the reasons for them. Other such questions might include, for example, if vaccine supplies are limited, who should get vaccinated? Can treating (Anderson 1991; Baggaley et al. 2006) or preventing (Panagiotopoulos et al. 1999) an infection ever be harmful? Why was smallpox the first disease to be eradicated by public health efforts? Which is harder to control, SARS or pandemic influenza, and why?

The first lecture should introduce key terminology and concepts relevant to infectious disease epidemiology, including such terms as reservoir, transmission route (airborne/droplet/waterborne/vector/etc.), communicable versus infectious, vector, host, etc. This is a good time to begin mentioning specific infections in connection with some of these concepts, in order to familiarize non-experts with common communicable diseases. For students needing introductions to the epidemiology of particular diseases, the twin volumes

edited by Evans and colleagues (Evans and Kaslow 1997; Evans and Brachman 1998) are an excellent resource.

The natural history of a communicable disease should be introduced here, with a clear distinction made between the natural history of infectiousness (latent period leading to infectious period) and the natural history of symptoms (incubation period leading to symptomatic period), making note of the fact that the incubation period may be either longer or shorter than the latent period. The concept of the incubation period should be distinguished from the notion of induction time commonly used in chronic disease epidemiology. Students may be asked at this point to consider the implications for disease control if the latent period is either longer or shorter than the infectious period, an aspect that will be formally addressed later. The serial interval or generation time is also a key concept here; its definition is inconsistent in the literature but a clear and self-consistent definition of the generation interval is the difference between the time at which an individual becomes infected and the time at which his/her infector became infected. A timeline will help to clarify these concepts (Giesecke 2001).

Following this survey of terminology, students should be introduced to the key features that make communicable diseases special and justify a distinct course on their epidemiology. These include (Giesecke 2001)

- the existence of immunity, both as a factor influencing risk and as a record of prior infection (Edmunds et al. 2000);
- the fact that a case may be a source and that, in some cases, a subclinical case may be a source;
- enormous and rapid variation in incidence and prevalence—the existence of epidemics;
- urgency—that relatively small delays in action may have large consequences (Ferguson et al. 2005);
- the existence of risk factors for transmission, not only for disease incidence or outcome (Quinn et al. 2000);
- the biological characteristics of the cause—the fact that pathogens have ecologic niches and can evolve (Levin et al. 1999);
- the fact that the impact of interventions in the population is not a simple sum of individual effects (e.g. indirect protection from vaccines);
- the fact that, statistically, outcomes in different persons are not independent (Cooper and Lipsitch 2004); and
- different causal structure—a single, by definition, 'necessary' cause versus 'complex' diseases with many sufficient causes.

Examples from the lecturer's own experience or field of study will be valuable in illustrating these points.

Trends in infectious disease should be discussed; key features are the decline in infectious disease mortality prior to the availability of biologically based treatment or prevention (Armstrong et al. 1999) (or even biological understanding; McKeown 1976), the notion of the epidemiological transition (Omran 1971), the decline in interest in infectious diseases, the persistence of infectious diseases in developing countries, and the resurgence of both disease and interest in disease with the appearance of HIV/AIDS.

Finally, the issue of causality in infectious disease should be discussed. A minimal approach would introduce Koch's postulates and give a single example, followed by some of the practical difficulties in fulfilling these postulates (unculturable pathogens, impossibility of truly isolating individual pathogens from particular contaminants, etc.). A fuller exploration could involve a whole lecture and could delve deeper into case studies such as the evidence for HIV as the cause of AIDS or the prion hypothesis for transmissible spongiform encephalopathies and could discuss the modern 'molecular Koch's postulates'. A brief account of the modern application of Koch's postulates for the SARS virus has been presented by Fouchier and colleagues (2003).

Epidemics in closed populations, and compartmental modelling as a framework for understanding (two to four hours)

This lecture should introduce the notion that an epidemic curve for a communicable disease in a closed population has a characteristic shape. Students should leave with a qualitative understanding of the reasons for each phase of the epidemic. Exponential growth is expected when a communicable disease is first introduced into a population, as each individual (on average) infects a fixed number of other individuals (the basic reproductive number—the term may be introduced here or deferred until its own lecture). This exponential growth begins to slow as the proportion of susceptible individuals among the contacts of infectious persons declines (since these contacts are now either themselves infected or immune); a peak is reached when each infection, on average, generates only one further infection, thus just replacing itself. This process continues to deplete susceptible hosts, further lowering the mean number of secondary cases per case; hence, the epidemic must decline. Some susceptibles are left over at the end of the epidemic, because disease declines not when the number of susceptibles approaches zero but when each case creates less than one secondary case on average. Analogies to wildfires spreading in a forest or to predator–prey dynamics in ecology may be helpful for students of appropriate backgrounds.

Following this qualitative introduction, a simple compartmental model should be introduced to quantify the ideas just expressed and to allow further analysis. The goal is to build up to a basic susceptible–infectious–recovered (SIR) model. Time should be taken in attempting to provide intuition for what the model represents. A two-compartment model for a non-communicable disease could be presented first, with fixed incidence and duration; this connects the compartmental modelling framework to the concepts of incidence, duration, and prevalence, all of which should be familiar after a basic epidemiology course. It also allows the instructor to make explicit the most fundamental aspect of a communicable disease—that incidence is not fixed or simply environmentally determined but depends on current prevalence. From here, the explicit form of an SIR model can be presented; clear introductions to this methodology are given in the works by Giesecke (2001), Anderson and Nokes (2002), and Keeling and Rohani (2007). Both box-arrow and differential equation forms may be presented.

Up to one to two hours may be spent in discussing the parameters of these models: contact rates, transmission probability, duration, and their measurement or estimation in real diseases. This discussion should emphasize the behavioural and biological assumptions underlying the models and the parameters (random mixing, contact rates dependent on densities of hosts, etc.).

The concepts of closed-community epidemics may be illustrated with data from well-described epidemics of smallpox (Eichner and Dietz 2003) or influenza (Keeling and Rohani 2007). An important application of these ideas in practice is given in the report by UNAIDS (1999).

At this point, a laboratory session may be used to introduce compartmental modelling by providing students with a basic SIR model and allowing them to explore its properties by changing duration, transmission probability, contact rate, or population density/size. For ease of learning and use, the Berkeley Madonna differential equation software is ideal. It is available as a free download trial version from <http://www.berkeleymadonna.com/>, and volume licences for students are available at reasonable prices. Students should be guided through the mechanics of manipulating an existing model at first so that the focus is on understanding transmission behaviour and its qualitative features rather than on understanding the models or software. More advanced students may wish to pursue programming within Berkeley Madonna to create more detailed models. Our infectious disease epidemiology is transitioning to teaching students to model in R (<http://www.r-project.org/>), which offers much greater flexibility and breadth of application, at the expense of more effort to acquire basic fluency.

Reproductive numbers, their implications, and their estimation (two hours)

The basic reproductive number R_0 is the most fundamental quantity in measuring the transmissibility of a communicable disease. The concept has been foreshadowed (or perhaps introduced) previously but should be defined here as the expected number of secondary infections caused by a typical infectious individual in a fully susceptible population. It is crucial to emphasize that this is determined partly by biological factors (e.g. pathogen shedding) and partly by social/behavioural ones (e.g. crowding and contact rates); thus, R_0 is not a constant of nature but a characteristic of a pathogen (or strain) in a given population. Representative values of R_0 should be presented (Anderson and May 1991), although these estimates are trustworthy only for the immunizing childhood infections (e.g. pertussis, measles, smallpox, mumps); population variability and uncertainty about how to estimate R_0 for many other diseases means that numbers should be treated with extreme scepticism for many other infections, including HIV, malaria, and many partially immunizing pathogens.

The relationship between R_0 and the effective reproductive number R_E should be stressed; the latter measures the actual number of secondary infections per case, given the presence of interventions, partially immune populations, etc. The concept of epidemic peak presented in 'Epidemics in closed populations, and compartmental modelling as a framework for understanding' should be revisited and formalized as the point where $R_E = 1$. The remainder of the lecture should be focused on the relationship between R_0 and other quantities of interest, including the final size of an epidemic in a closed population, the probability that an epidemic takes off, and the growth rate of an epidemic in its early phase.

Laboratory work associated with this session should include changing contact rates, transmission probabilities, and durations to alter the basic reproductive number, allowing students to make predictions about the impact of such changes on the ability of the infection to spread and its rate of spread.

Herd immunity and control of a communicable disease (two hours)

This lecture should introduce the concept of herd immunity, again reinforcing the notion that what happens to one individual affects other individuals in the population. The notion of herd immunity—the existence of a level of immunity in the population that affects transmission to those who are not immune—may be introduced side by side with the notion of indirect protection: that immunizing one individual by natural exposure or vaccination may protect another individual. These two concepts are distinct and should be introduced as such;

however, in practice, they are used interchangeably, even by distinguished epidemiologists.

Students should be challenged to consider (1) for which kinds of diseases and which kinds of intervention one might observe indirect protection, and (2) what sorts of evidence might be gathered to test whether indirect protection results from a vaccine. Data including reductions in infectiousness in vaccinated persons, reductions in disease incidence by a fraction greater than the fraction vaccinated, reductions in disease incidence in household members of vaccinees, or reductions in age groups outside the target ages of vaccination should be discussed. Fine (1993) has an excellent review that also addresses dynamical changes expected with indirect protection, for example, changes in the mean age of infection or in the frequency of epidemics. With advanced students, this article provides a good introduction to topics that will be covered later. Another advantage of this article is its review of herd immunity in particular diseases.

Herd immunity provides a context in which to discuss why elimination or eradication of a disease is possible with immunization of less than 100 per cent of the population. The equation for the critical fraction to immunize to eliminate transmission, $P_c = 1 - 1/R_0$, should be derived from the consideration of R_0 and R_E and introduced at this stage. This could lead to a discussion of the factors affecting the ease of eradicating a disease—low R_0, effective vaccine, no non-human reservoir—and the reason why smallpox was the first disease eradicated by vaccination.

This list, however, is incomplete, even in the case of smallpox, since smallpox eradication did not simply depend on mass vaccination but also on 'mop-up' operations via 'ring vaccination' of contacts of cases. A discussion of the history of smallpox vaccination may be of interest here (Fenner et al. 1988) and the conceptual issues to be raised include the additional importance of disease natural history and the timing of control measures: smallpox ring vaccination was possible because the latent period is nearly as long as, or perhaps as long as, the incubation period, so that identifying symptomatic persons was a reliable guide to identifying potentially infected contacts. Moreover, those contacts could be vaccinated after exposure and still protected (Fenner et al. 1988). A good conceptual discussion of the issues affecting the success of control measures in general is given by Fraser et al. (2004).

Transmission dynamics in open populations with births and deaths (one to two hours)

The goal of this session is to answer one of the questions posed at the opening of the course: the origin of repeated, annual, or multiannual epidemics of communicable diseases. An intuitive way to introduce the problem is to compare

the epidemic in a closed community (unimodal, ending with zero cases) qualitatively against data from pre-vaccination childhood diseases such as measles (non-zero cases, recurrent epidemics, sometimes annually and sometimes in cycles of two or more years). A model implemented in Berkeley Madonna may be set up in class to begin from a closed community and slowly add 'features' (give non-zero values to parameters such as birth rates and seasonal fluctuations) to produce dynamics with increasing similarity to the ones observed in data. Given that an epidemic dies out due to loss of susceptibles, the birth of new susceptibles is clearly required to maintain transmission over time. Adding births to a simple model produces an endemic equilibrium level of infection without oscillations or with damped oscillations, depending on parameter values. Persistent oscillations may be obtained by adding seasonal forcing to the model via a sinusoidal term (Keeling and Rohani 2007). The period of the oscillations is predictable from the birth rate, the serial interval, and the basic reproductive number (Anderson and May 1991). If students have computers available in the classroom, they may be able to follow along with this demonstration using Berkeley Madonna; alternatively, some of this material may be taught in a laboratory session. More advanced classes may wish to cover the material/application described by Earn and colleagues (2000), who used changing patterns of birth rates and vaccination to explain changes in the patterns of epidemics of measles over time.

Laboratory work should allow students to work with the full model of an infection with seasonal forcing and births. They should be asked to predict the period of oscillations for various parameters and to check it against the model runs. This is also a good time to reinforce earlier concepts about the supply of susceptibles as driving the epidemic.

Age-seroprevalence studies and inferences from them (two to three hours)

The purpose of this session is to introduce the notion of cross-sectional data on the prevalence of immunity in different age groups, and the use of that data (under particular assumptions) to make inferences about the transmission of the disease. Biological aspects of immunity may be covered here, at a level of detail appropriate for the audience. Key points are as follows:

For some infections, acquired immunity is virtually complete and lifelong (e.g. for measles), while for others it may be partial, temporary and/or strain-specific (e.g. influenza, malaria), or virtually non-existent (e.g. gonorrhoea).

- For infections with lasting and highly effective immunity, individuals may be classified into susceptible or immune, usually based on their antibody

levels, and the prevalence of immunes is called the seroprevalence. A selection of age-seroprevalence curves (seroprevalence by age group) has been presented by Anderson and May (1985).

- If a disease has been endemic in a population for a long time, then the cross-sectional pattern of immunity may reflect the experience of a cohort (i.e. seroprevalence should increase with age at a rate equal to the annual risk of infection).
- If the disease has not been endemic over a long period, cross-sectional age-seroprevalence may depart dramatically from the experience of a cohort; for example, a disease whose transmission has declined recently will show elevated and increasing seroprevalence in older persons but this does not indicate that current transmission is concentrated in older age groups.

Under the assumption that a disease has long been endemic in a population, it is possible to calculate age-specific forces of infection from age-seroprevalence data. Students should be walked through an example of this process using real or simulated data; it may be useful to have them design an Excel spreadsheet to perform the calculations. Laboratory sessions should give the students practice doing these calculations.

This is also an appropriate stage to introduce the simple relationship among the basic reproductive number $R_{0,}$ the mean age at infection A, and the lifespan L, for a disease with lifelong immunity spreading in a homogeneously mixed population: $R_0 \approx L/A$. The calculation may be justified by a graphical or intuitive argument (Anderson and May 1991). It should be emphasized that this calculation can be strongly misleading when the assumptions are violated (Gupta et al. 1994).

Heterogeneity in infectious disease transmission (two to four hours)

For every communicable disease, hosts differ in their biological susceptibility, their biological tendency to be infectious, and their behavioural proclivity to make infectious contacts. These heterogeneities may affect every aspect of infectious disease transmission, including the probability that a single case will spark an outbreak, the rate at which a disease initially spreads, the eventual size of an epidemic, and the effectiveness and optimal targeting of control measures. Classically, this heterogeneity has been best studied as it relates to age (mainly for directly transmitted diseases) and to sexual activity levels (for sexually transmitted infections).

Students may be introduced to the issue of heterogeneity through examples including superspreading in SARS and other diseases (Shen et al. 2004;

Lloyd-Smith et al. 2005), the importance of children in spreading respiratory diseases (Reichert et al. 2004; Lexau et al. 2005), and the effect of viral load on infectiousness in HIV (Quinn et al. 2000). A key objective in this lecture is to help students to differentiate among the types of heterogeneity and understand their effects. The contribution of an individual to disease transmission depends on three key factors:

1. the frequency of potentially infectious contacts, which in general affects both the probability that the individual becomes infected and the number of secondary infections that individual may cause—this in turn may depend on the individual's degree of hygiene, degree of crowding in the home, use of precautions such as condoms for a sexually transmitted disease, and professional and personal habits that affect contact rates;
2. the probability of becoming infected upon exposure—this may depend on such factors as the individual's immune status (history of infection with the pathogen in question as well as general immune sufficiency), whether the individual is circumcised (for certain sexually transmitted diseases), the individual's hygiene or use of precautions, etc.; and
3. the individual's infectiousness to others once infected; this again may depend on hygiene, immune status, circumcision, site of infection, degree of illness, or other factors.

These proximate causes of heterogeneity are often unknown for particular persons in the population but easily observable characteristics such as age, sex, or profession may act as proxies for these causes and may be useful ways of classifying persons for the purpose of understanding their contribution to transmission.

Age is perhaps the simplest and most intuitively straightforward axis of heterogeneity with which to introduce the quantitative analysis of the issue. The section 'Age-seroprevalence studies and inferences from them' introduced the estimation of age-specific force of infection from seroprevalence data. For a basic class, one may wish simply to discuss the issue of age-heterogeneity in infectiousness and susceptibility in general terms here. For a more advanced class, one could go into greater detail on the use of assumed mixing matrices (who-acquires-infection-from-whom, or WAIFW, matrices) to estimate transmission from the age-specific force of infection and inferred prevalence and incidence. There is no simple treatment of this process in existing textbooks but the exposition in the book by Anderson and May (1991) is quite complete. A key limitation of this approach is that, with g groups, there are g^2 mixing coefficients but only g constraints, so it is necessary when estimating group-to-group transmission coefficients to assume

equality (or at least dependence) among some of these mixing rates. This approach has always been somewhat unsatisfactory but recent efforts to estimate social mixing patterns directly (Wallinga et al. 2006) may begin to provide viable alternative approaches.

Applications of age-related heterogeneity in transmission to the design of control measures include pulsed vaccination strategies (Keeling and Rohani 2007), vaccination of children to protect adults (Lexau et al. 2005; Halloran and Longini 2006), and design of policies to avoid increases in disease incidence at vulnerable ages (e.g. for rubella (Anderson and May 1991) and varicella (Edmunds and Brisson 2002)).

Heterogeneity in sexually transmitted infections (one hour or more)

Few features of human behaviour are as variable as the frequency of acquiring new sexual partners, which ranges from celibacy or mutual monogamy at one extreme (while the subjective difference between these is vast, they are equally disadvantageous to the transmission of a sexually transmitted infection (STI)!), to extreme promiscuity at the other. An enormous literature assesses the theoretical impact of such variability on the dynamics of STIs, yet the empirical impact of these heterogeneities remains unclear because of the imprecision of our measurements of human behaviour.

At least three aspects of heterogeneity should be presented and their effects distinguished in the context of STIs:

1. Variation in the rate of partner acquisition per se. Simply the fact that some individuals have more sexual partnerships than others tends to facilitate and speed up the spread of STIs in a population, compared to a situation in which everyone has the mean number of sexual partners. The reason for this is that individuals who have many partners are both more likely to become infected and more likely to infect a large number of others if they do become infected. The effect here is to make the basic reproductive number proportional to the mean rate of partner acquisition plus the ratio of the variance to the mean (Anderson and May 1991). For quantitatively advanced classes, it may be interesting to consider highly counterintuitive consequences of this idea (Kremer and Morcom 1998) and to discuss whether the theoretical assumptions can support these consequences.
2. Assortativeness in partner choice. Conceptually, assortativeness refers to a tendency of individuals who have many sexual partners to choose as partners others who also have many sexual partners. This may be contrasted with proportionate or 'random' mixing, in which individuals choose

partners without regard to whether the partner has a similar activity level (though highly active individuals will be over-represented in the pool of potential partners for everyone, simply because they are highly active). Assortativeness may also be contrasted with disassortative mixing, in which individuals tend to partner with persons with dissimilar activity levels (e.g. commercial sex-workers with individuals who have few partners). In general, assortativeness tends to increase the likelihood that an infection can persist in a population (since once it enters a high-risk group, it tends to continue circulating in that group) but may in some cases reduce the prevalence of the infection (since contacts of highly infectious people may be 'wasted' on others who are likely already to be infected or immune; Garnett and Anderson 1996).

3. Partner concurrency. In recent years, it has become clear that concurrent sexual partnerships are a major factor allowing rapid spread of STIs—especially HIV, for which the first months of infection carry the greatest transmission risk (Gorbach and Holmes 2003). For this reason, a concept of STI epidemiology as resulting from sequential sexual partnerships is incomplete and may underestimate the potential for spread (Morris and Kretzschmar 1997).

The concept of a 'core group' (Yorke et al. 1978) that is responsible for maintenance of an STI in a population is central to much thinking in STI epidemiology.

Emerging diseases (one hour or more)

The appearance of new diseases in highly susceptible populations poses a number of epidemiologic problems, as well as problems of control. Most novel human diseases emerge as zoonoses, of which a subset acquires the ability to transmit within human populations. Long lists of the determinants of such disease emergence events have been compiled (Morse 1995) but a long list of factors is an expression of uncertainty rather than understanding. There have been efforts to quantify the contributions of various factors to disease emergence theoretically (Antia et al. 2003) and empirically (Jones et al. 2008).

Once a disease acquires the ability to transmit within human populations, several questions arise. The first is whether a disease is likely to spread successfully in a particular population; the dramatic differences between countries in the degree of the spread of SARS in 2003 provide a dramatic illustration. The most important determinant of spread is the basic reproductive number R_0. If it is less than 1, then a large epidemic is not possible while, if it exceeds 1, then an epidemic is possible but not guaranteed; the probability increases as a function

of R_0. However, R_0 represents a mean, and the degree of variation around that mean also determines the probability that the epidemic will 'take off'. Accounts of the basic theory applied to particular diseases are given by Lipsitch and colleagues (2003) and Lloyd-Smith and colleagues (2005).

A second question is how fast the epidemic will spread. Spread will be more rapid for larger R_0 and slower for longer serial intervals. For a simple SIR-type infection, the exponential growth of the early epidemic will occur at a rate given by $(R_0 - 1)/D$, where D is the duration of infectiousness or, equivalently in this model, the serial interval; this concept may have been described in the lecture on R_0. Theory for more complex natural histories has been reviewed by Wallinga and Lipsitch (2007). Turning this around, the rate of change in case numbers, combined with information about the distribution of serial intervals, may be used to estimate the current effective reproductive number with some time lag; the key reference for this topic is the paper by Wallinga and Teunis (2004).

The estimation of case–fatality proportion during an emerging epidemic may also be addressed if time permits (Ghani et al. 2005).

The foregoing constitutes a basic course in the concepts of communicable disease epidemiology. Many aspects of the topic are of course not covered. A more extensive course might include modules on some of the following topics.

Vector-borne diseases

Some of the first developments of the concepts of infectious disease dynamics were motivated by problems of malaria, a vector-borne disease (Ross 1916, 1917; Ross and Hudson 1917). Aspects of vector ecology and behaviour affect many of the basic assumptions about transmission dynamics and may justify a separate session on this class of diseases. A remarkable special issue of the *American Journal of Tropical Medicine and Hygiene* containing a collection of papers on a trial of insecticide-treated bed nets for malaria provides a rich source of epidemiologic concepts and their applications to this key vector-borne disease (Nahlen et al. 2003).

Study design for communicable diseases

Communicability introduces a number of complexities into study design. In particular, the non-independence of outcomes means that studies of interventions may miss indirect effects or may be biased by dependencies between patients. The design of studies to capture indirect effects while avoiding bias is an active area of research; indeed, a whole course in this area could be taught. Aspects that might be addressed in one or more lectures include

- randomized trials—the inability of standard, individual-randomized trials to detect population-level effects, and the use of community-randomized designs to capture indirect effects at the community level (Moulton et al. 2001; Halloran and Longini 2006);
- the use of household or partner studies to estimate effects of treatment or vaccination on infectiousness (Datta et al. 1998; Halloran et al. 2007);
- the use of ecological studies to infer indirect protection from the vaccination of selected age groups (Reichert et al. 2001; Lexau et al. 2005);
- the use of vaccine-probe studies to estimate burden of disease attributable to a particular pathogen (Gessner et al. 2005; Scott 2007); and
- the problems of non-independent outcome in studies of hospital-acquired infections and the use of appropriate methods (Cooper and Lipsitch 2004).

Many, though not all, of these topics are discussed by Halloran (2001).

Evolution of virulence and antigenic diversity in pathogens

The biological interaction between host and pathogen is a distinctive feature of communicable diseases and raises the possibility of pathogen evolution in response to selection imposed by the host. A classical notion that host–pathogen interactions always evolve toward benignness has been partly replaced by an understanding that harming the host may provide a transmission advantage to the pathogen (by prolonging infection, delaying clearance by the immune system, and increasing the production of infectious propagules) and that this advantage may offset the 'cost' to the pathogen of killing or disabling its current host (Anderson and May 1991). Classical experimental results with myxomatosis in rabbits should be covered in any lecture on this subject (Fenner and Ratcliffe 1965; Anderson and May 1991); unfortunately, there have been few empirical follow-ups to these studies. A large theoretical literature has developed on the 'evolution of virulence', yet biological differences among infections frustrate attempts to generalize (Ebert and Bull 2003). This approach to studying virulence from the perspective of selection may be contrasted with the rather different molecular genetics approach of studying 'virulence determinants', defined as genes in a pathogen that are essential to cause disease or to infect an animal host but not for growth in culture (Finlay and Falkow 1997).

The host immune system also exerts selection pressure on pathogen populations to diversify in order to escape host immunity; this, in turn, alters the pattern of disease to which hosts are exposed. Examples at a range of temporal and spatial scales include within-host escape from cytotoxic T cell immunity in

HIV (Rambaut et al. 2004), the diversification of influenza A virus hemagglutinin (Koelle et al. 2006), and the population-wide diversity of serogroups/serotypes in colonizing bacteria (Lipsitch and O'Hagan 2007).

Evolution in response to infectious disease interventions

The evolution of pathogens is driven not only by biological interactions with the host but also by our deliberate interventions for prevention and treatment. Antimicrobial resistance is the most important example, another topic that could be the subject of a lecture or a whole course. Some biological background is very helpful to set the stage, including at least the genetic mechanisms of resistance (mutation, acquisition of foreign DNA, including plasmids) and biological mechanisms of resistance (e.g. enzymatic destruction, target change or amplification, efflux, etc.). From an epidemiological perspective, there is a key distinction between the risk of emergence of resistance during treatment (particularly worrisome for mutationally conferred resistance, e.g. in tuberculosis or HIV infection) and the use of anti-infectives that select for spread of resistance at the population level (Lipsitch and Samore 2002). Resistance raises special issues of study design (Lipsitch 2001a; A. Harris et al. 2002) and mathematical modelling (Lipsitch 2001b; Bootsma et al. 2006).

Vaccination, too, imposes selective pressures on pathogen populations, in some cases leading to the selection of variants that escape from the vaccine. A clear example is the emergence of non-vaccine-type disease in *Streptococcus pneumoniae* (Hicks et al. 2007); the significance of this phenomenon is less clear in pertussis (Mooi et al. 2001) and hepatitis B (Wilson et al. 2000). In the case of oral polio vaccine, it is the vaccine virus itself whose evolution poses a challenge to eradication efforts (Kew 2005).

Role of genomics in modern infectious disease epidemiology

With the advent of affordable genome sequencing for pathogens, whole genomes are increasingly a key part of the evidence brought to bear on infectious disease outbreaks (Fraser et al. 2009; Grad et al. 2012; S. Harris et al. 2013) and endemic infectious diseases (Mutreja et al. 2011; Croucher et al. 2013). Some of this activity is for specialized evolutionary studies but I speculate that it is only a matter of a few years before 'pathogen genome sequence(s)' will become a standard element of epidemiologic line lists, along with elements of person, place, and time. Methods to use these techniques in outbreak investigations are in their infancy but extra time could profitably be devoted to exposing students to early examples of this growing trend.

Miscellaneous topics in infectious disease epidemiology

Other topics worthy of consideration that may be covered according to instructor and audience interest include

- the impact of host genetics on patterns of infectious disease incidence (Hill 2006; Casanova and Abel 2007);
- dose–response relationships in infectious diseases (including non-communicable diseases; Haas et al. 1999);
- surveillance—traditional and syndromic (Brookmeyer and Stroup 2003; Lawson and Kleinman 2005); and
- spatial spread of communicable diseases (Cliff and Haggett 1992; Keeling and Rohani 2007).

Assessing students' achievements

Student achievement may be assessed by performance on problem sets and/or computer laboratory assignments throughout the course and through exams at the midpoint and end of the course.

This course is taught at a fairly high level of abstraction. Students should have an opportunity to see the application of these concepts to real problems. Critiques and discussions of research papers are an excellent way to provide such exposure while also evaluating student understanding. Depending on time available, such critiques may be in the form of group discussion, individual oral presentation, or written student papers. Some papers that could spark good discussion include

- a study of individual and group-level effects of vaccine exemption (Feikin et al. 2000);
- some of the papers on the direct and indirect effects of insecticide-treated bed nets for malaria control collected in a special journal supplement (Nahlen et al. 2003)—the paper by Hawley and colleagues (2003) is a good choice;
- the papers cited in the sections entitled 'Introduction, terminology, trends, techniques, and causality' and 'Heterogeneity in infectious disease transmission' on the possible perverse effects of varicella or rubella vaccination or of HIV treatment or prevention;
- the UNAIDS (1999) report cited in the section titled 'Epidemics in closed populations and compartmental modelling as a framework for understanding' on the interpretation of falling HIV incidence and whether this can be attributed simply to the natural course of the epidemic;

- the papers on the relationship between viral load and infectiousness in HIV (Quinn et al. 2000; Gray et al. 2001); and
- a selection of papers discussing responses to a smallpox bioterror event (a good synthesis is given in Ferguson et al. 2003), an influenza pandemic (Ferguson et al. 2005, 2006; Germann et al. 2006; Mills et al. 2006), or the SARS epidemic (Lipsitch et al. 2003; Riley et al. 2003; Wallinga and Teunis 2004).

References

Anderson, R. M., Gupta, S., and May, R. M. (1991) Potential of community-wide chemotherapy or immunotherapy to control the spread of HIV-1. *Nature*, **350**: 356–9.

Anderson, R. M. and May, R. M. (1985) Vaccination and herd immunity to infectious diseases. *Nature*, **318**: 323–9.

Anderson, R. M. and May, R. M. (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press.

Anderson, R. M. and Nokes, D. J. (2002) 'Mathematical models of transmission and control', in R. Detels, J. **McEwen**, R. Beaglehole, and H. Tanaka, eds, *Oxford Textbook of Public Health* (4th edn). New York: Oxford University Press, pp. 715–44.

Antia, R., Regoes, R. R., Koella, J. C., and Bergstrom, C. T. (2003) The role of evolution in the emergence of infectious diseases. *Nature*, **426**: 658–61.

Armstrong, G. L., Conn, L. A., and Pinner, R. W. (1999) Trends in infectious disease mortality in the United States during the 20th century. *Journal of the American Medical Association*, **281**: 61–6.

Baggaley, R. F., Garnett, G. P., and Ferguson, N. M. (2006) Modelling the impact of antiretroviral use in resource-poor settings. *PLoS Medicine*, **3**: e124.

Bootsma, M. C., Diekmann, O., and Bonten, M. J. (2006) Controlling methicillin-resistant *Staphylococcus aureus*: quantifying the effects of interventions and rapid diagnostic testing. *Proceedings of the National Academy of Sciences USA*, **103**: 5620–5.

Brookmeyer, R. and Stroup, D. (2003) *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*. New York: Oxford University Press.

Casanova, J. L. and Abel, L. (2007) Human genetics of infectious diseases: a unified theory. *European Molecular Biology Organization Journal*, **26**: 915–22.

Cliff, A. D. and Haggett, P. (1992) *Atlas of Disease Distributions: Analytical Approaches to Epidemiological Data*. Oxford: Basil Blackwell.

Cooper, B. and Lipsitch, M. (2004) The analysis of hospital infection data using hidden Markov models. *Biostatistics*, **5**: 223–37.

Croucher, N. J., Finkelstein, J. A., Pelton, S. I., Mitchell, P. K., Lee, G. M., Parkhill, J., Bentley, S. D., Hanage, W. P., and Lipsitch, M. (2013) Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nature Genetics*, **45**: 656–63.

Datta, S., Halloran, M. E., and Longini, I. M. Jr. (1998) Augmented HIV vaccine trial design for estimating reduction in infectiousness and protective efficacy. *Statistics in Medicine*, **17**: 185–200.

Diekmann, O. and Heesterbeek, J. A. P. (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. New York: John Wiley & Sons.

Earn, D. J., Rohani, P., Bolker, B. M., and Grenfell, B. T. (2000) A simple model for complex dynamical transitions in epidemics. *Science*, **287**: 667–70.

Ebert, D. and Bull, J. J. (2003) Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends in Microbiology*, **11**: 15–20.

Edmunds, W. J. and Brisson, M. (2002) The effect of vaccination on the epidemiology of *varicella zoster* virus. *Journal of Infectious Diseases*, **44**: 211–19.

Edmunds, W. J., Gay, N. J., Kretzschmar, M., Pebody, R. G., and Wachmann, H., on behalf of the ESEN Project, European Sero-Epidemiology Network. (2000) The pre-vaccination epidemiology of measles, mumps and rubella in Europe: implications for modelling studies. *Epidemiology and Infection*, **125**: 635–50.

Eichner, M. and Dietz, K. (2003) Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, **158**: 110–17.

Evans, A. S. and Brachman, P. S., eds. (1998) *Bacterial Infections of Humans: Epidemiology and Control* (3rd edn). New York: Springer.

Evans, A. S. and Kaslow, R. A., eds. (1997) *Viral Infections of Humans: Epidemiology and Control*. New York: Springer.

Feikin, D. R., Lezotte, D. C., Hamman, R. F., Salmon, D. A., Chen, R. T., and Hoffman, R. E. (2000) Individual and community risks of measles and pertussis associated with personal exemptions to immunization. *Journal of the American Medical Association*, **284**: 3145–50.

Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., Ladnyi, I. D., and WHO. (1988) *Smallpox and its Eradication*. Geneva: WHO.

Fenner, F. and Ratcliffe, F. N. (1965) *Myxomatosis*. Cambridge: Cambridge University Press.

Ferguson, N. M., Cummings, D. A., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., and Burke, D. S. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, **437**: 209–14.

Ferguson, N. M., Cummings, D. A., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. (2006) Strategies for mitigating an influenza pandemic. *Nature*, **442**: 448–52.

Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., Anderson, R. M., and Leach, S. (2003) Planning for smallpox outbreaks. *Nature*, **425**: 681–5.

Fine, P. E. M. (1993) Herd immunity: history, theory, practice. *Epidemiologic Reviews*, **15**: 265–302.

Finlay, B. B. and Falkow, S. (1997) Common themes in microbial pathogenicity revisited. *Microbiology and Molecular Biology Reviews*, **61**: 136–69.

Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G., Peiris, M., Lim, W., Stöhr, K., and Osterhaus, A. D. (2003) Aetiology: Koch's postulates fulfilled for SARS virus. *Nature*, **423**: 240.

Fraser, C. et al. (2009) Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*, **324**: 1557–61.

Fraser, C., Riley, S., Anderson, R. M., and Ferguson, N. M. (2004) Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences USA*, **101**: 6146–51.

Garnett, G. P. and Anderson, R. M. (1996) Sexually transmitted diseases and sexual behavior: insights from mathematical models. *Journal of Infectious Diseases*, **174 Suppl. 2**: S150–61.

Germann, T. C., Kadau, K., Longini, I. M. Jr, and Macken, C. A. (2006) Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences USA*, **103**: 5935–40.

Gessner, B. D. et al. (2005) Incidences of vaccine-preventable Haemophilus influenzae type b pneumonia and meningitis in Indonesian children: hamlet-randomised vaccine-probe trial. *Lancet*, **365**: 43–52.

Ghani, A. C., Donnelly, C. A., Cox, D. R., Griffin, J. T., Fraser, C., Lam, T. H., Ho, L. M., Chan, W. S., Anderson, R. M., Hedley, A. J., and Leung, G. M. (2005) Methods for estimating the case fatality ratio for a novel, emerging infectious disease. *American Journal of Epidemiology*, **162**: 479–86.

Giesecke, J. (2001) *Modern Infectious Disease Epidemiology* (2nd edn). New York: Oxford University Press.

Gorbach, P. M. and Holmes, K. K. (2003) Transmission of STIs/HIV at the partnership level: beyond individual-level analyses. *Journal of Urban Health*, **80 Suppl. 3**: iii15–25.

Grad, Y. H. et al. (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proceedings of the National Academy of Sciences USA*, **109**: 3065–70.

Gray, R. H. et al. (2001) Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1–discordant couples in Rakai, Uganda. *Lancet*, **357**: 1149–53.

Gupta, S., Trenholme, K., Anderson, R. M., and Day, K. P. (1994) Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. *Science*, **263**: 961–3.

Haas, C. N., Rose, J. B., and Gerba, C. P. (1999) *Quantitative Microbial Risk Assessment*. New York: John Wiley & Sons.

Halloran, M. E. (2001) 'Overview of study design', in J. C. Thomas and D. J. Weber, eds, *Epidemiologic Methods for the Study of Infectious Diseases*. New York: Oxford University Press, pp. 86–115.

Halloran, M. E., Hayden, F. G., Yang, Y., Longini, I. M. Jr, and Monto, A. S. (2007) Antiviral effects on influenza viral transmission and pathogenicity: observations from household-based trials. *American Journal of Epidemiology*, **165**: 212–21.

Halloran, M. E. and Longini, L. M. Jr. (2006) Public health: community studies for vaccinating schoolchildren against influenza. *Science*, **311**: 615–16.

Harris, A. D., Samore, M. H., Lipsitch, M., Kaye, K. S., Perencevich, E., and Carmeli, Y. (2002) Control-group selection importance in studies of antimicrobial resistance: examples applied to *Pseudomonas aeruginosa*, enterococci, and *Escherichia coli*. *Clinical Infectious Diseases*, **34**: 1558–63.

Harris, S. R. et al. (2013) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infectious Diseases*, **13**: 130–6.

Hawley, W. A. et al. (2003) Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. *American Journal of Tropical Medicine and Hygiene*, **68 Suppl.**: 121–7.

Hicks, L. A. et al. (2007) Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998–2004. *Journal of Infectious Diseases*, **196**: 1346–54.

Hill, A. V. (2006) Aspects of genetic susceptibility to human infectious diseases. *Annual Review of Genetics*, **40**: 469–86.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature*, **451**: 990–3.

Keeling, M. J. and Rohani, P. (2007) *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press.

Kew, O. M., Sutter, R.W., de Gourville, E. M., Dowdle, W. R., and Pallansch, M. A. (2005) Vaccine-derived polioviruses and the endgame strategy for global polio eradication. *Annual Review of Microbiology*, **59**: 587–635.

Knobler, S. L., Lemon, S. M., and Knobler, S. L., eds. (2004) *The Infectious Etiology of Chronic Diseases: Defining the Relationship, Enhancing the Research, and Mitigating the Effects*. Washington, DC: National Academies Press.

Koelle, K., Cobey, S., Grenfell, B., and Pascual, M. (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science*, **314**: 1898–903.

Kremer, M. and Morcom, C. (1998) The effect of changing sexual activity on HIV prevalence. *Mathematical Biosciences*, **151**: 99–122.

Lawson, A. B. and Kleinman, K. (2005) *Spatial and Syndromic Surveillance for Public Health*. New York: Wiley.

Levin, B. R., Lipsitch, M., and Bonhoeffer, S. (1999) Population biology, evolution, and infectious disease: convergence and synthesis. *Science*, **283**: 806–9.

Lexau, C. A., Lynfield, R. et al. (2005) Changing epidemiology of invasive pneumococcal disease among older adults in the era of pediatric pneumococcal conjugate vaccine. *Journal of the American Medical Association*, **294**: 2043–51.

Lipsitch, M. (2001a) Measuring and interpreting associations between antibiotic use and penicillin resistance in *Streptococcus pneumoniae. Clinical Infectious Diseases*, **32**: 1044–54.

Lipsitch, M. (2001b) The rise and fall of antimicrobial resistance. *Trends in Microbiology*, **9**: 438–44.

Lipsitch, M. et al. (2003) Transmission dynamics and control of severe acute respiratory syndrome. *Science*, **300**: 1966–70.

Lipsitch, M. and O'Hagan, J. J. (2007) Patterns of antigenic diversity and the mechanisms that maintain them. *Journal of the Royal Society Interface*, **4**: 787–802.

Lipsitch, M. and Samore, M. H. (2002) Antimicrobial use and antimicrobial resistance: a population perspective. *Emerging Infectious Diseases*, **8**: 347–54.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. (2005) Superspreading and the effect of individual variation on disease emergence. *Nature*, **438**: 355347–9.

Mazur, E. (2005) 'Qualitative versus quantitative reasoning: are we teaching the right thing?' in N. Sanitt, ed., *Motivating Science: Science Communication from a Philosophical, Educational and Cultural Perspective*. Luton: Panteneto Press, pp. 139–41.

McKeown, T. (1976) *The Role of Medicine: Dream, Mirage or Nemesis?* London: Nuffield Provincial Hospitals Trust.

Mills, C. E., Robins, J. M., Bergstrom, C. T., and Lipsitch, M. (2006) Pandemic influenza: risk of multiple introductions and the need to prepare for them. *PLoS Medicine*, **3**: e135.

Mooi, F. R., van Loo, I. H., and King, A. J. (2001) Adaptation of *Bordetella pertussis* to vaccination: a cause for its reemergence? *Emerging Infectious Diseases*, **7 Suppl.**: 526–8.

Morris, M. and Kretzschmar, M. (1997) Concurrent partnerships and the spread of HIV. *AIDS*, **11**: 641–8.

Morse, S. S. (1995) Factors in the emergence of infectious diseases. *Emerging Infectious Diseases*, **1**: 7–15.

Moulton, L. H., O'Brien, K. L., Kohberger, R., Chang, I., Reid, R., Weatherholtz, R., Hackell, J. G., Siber, G. R., and Santosham, M. (2001) Design of a group-randomized *Streptococcus pneumoniae* vaccine trial. *Controlled Clinical Trials*, **22**: 438–52.

Mutreja, A. et al. (2011) Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*, **477**: 462–5.

Nahlen, B. L., Clark, J. P., and Alnwick, D. (2003) Insecticide-treated bed nets. *American Journal of Tropical Medicine and Hygiene*, **68 Suppl.**: 1–2.

Nelson, K. E. and Masters Williams, C. (2006) *Infectious Disease Epidemiology: Theory and Practice* (2nd edn). Sudbury, MA: Jones & Bartlett Publishers.

Omran, A. R. (1971) The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Quarterly*, **49**: 509–38.

Panagiotopoulos, T., Antoniadou, I., and Valassi-Adam, E. (1999) Increase in congenital rubella occurrence after immunisation in Greece: retrospective survey and systematic review. *British Medical Journal*, **319**: 1462–7.

Quinn, T. C., Wawer, M. J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M. O., Lutalo, T., and Gray, R. H. (2000) Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *New England Journal of Medicine*, **342**: 921–9.

Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004) The causes and consequences of HIV evolution. *Nature Reviews Genetics*, **5**: 52–61.

Reichert, T. A., Simonsen, L., Sharma, A., Pardo, S. A., Fedson, D. S., and Miller, M. A. (2004) Influenza and the winter increase in mortality in the United States, 1959–1999. *American Journal of Epidemiology*, **160**: 492–502.

Reichert, T. A., Sugaya, N., Fedson, D. S., Glezen, W. P., Simonsen, L., and Tashiro, M. (2001) The Japanese experience with vaccinating schoolchildren against influenza. *New England Journal of Medicine*, **344**: 889–96.

Riley, S. et al. (2003) Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science*, **300**: 1961–6.

Ross, R. (1916) An application of the theory of probabilities to the study of a priori pathometry, Part I. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **92**: 204–30.

Ross, R. (1917) An application of the theory of probabilities to the study of a priori pathometry, Part II. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **93**: 212–25.

Ross, R. and Hudson, H. P. (1917) An application of the theory of probabilities to the study of a priori pathometry, Part III. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **93**: 225–40.

Scott, J. A. (2007) The preventable burden of pneumococcal disease in the developing world. *Vaccine*, **25**: 2398–405.

Shen, Z., Ning, F., Zhou, W., He, X., Lin, C., Chin, D. P., Zhu, Z., and Schuchat, A. (2004) Superspreading SARS events, Beijing, 2003. *Emerging Infectious Diseases*, **10**: 256–60.

Thomas, J. C. and Weber, D. J., eds. (2001) *Epidemiologic Methods for the Study of Infectious Diseases*. New York: Oxford University Press.

UNAIDS. (1999) *Trends in HIV Incidence and Prevalence: Natural Course of the Epidemic or Results of Behavioural Change?* Geneva: UNAIDS.

Vynnycky, E. and White, R. (2010) *An Introduction to Infectious Disease Modelling*. New York, Oxford University Press.

Wallinga, J., Kretzschmar, M., and Teunis, P. (2006) Using social contact data to estimate age-specific transmission parameters for infectious respiratory spread agents. *American Journal of Epidemiology*, **164**: 936–44.

Wallinga, J. and Lipsitch, M. (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, **274**: 599–604.

Wallinga, J. and Teunis, P. (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, **160**: 509–16.

Wilson, J. N., Nokes, D. J., and Carman, W. F. (2000) Predictions of the emergence of vaccine-resistant hepatitis B in The Gambia using a mathematical model. *Epidemiology and Infection*, **124**: 295–307.

Yorke, J. A., Hethcote, H. W., and Nold, A. (1978) Dynamics and control of the transmission of gonorrhea. *Sexually Transmitted Diseases*, **5**: 51–6.

Chapter 17

Cancer epidemiology

Pagona Lagiou and Dimitrios Trichopoulos

Introduction to cancer epidemiology

Cancer epidemiology is an important topic for health professionals in clinical medicine as well as public health. Teaching this subject, however, presents considerable difficulties, some of which are related to the background and expectations of the students, and others inherent to the topic of study itself. Some students lack in-depth knowledge of biological processes that are essential for the understanding of carcinogenesis and eventually cancer epidemiology. Other students, however, know more about biological processes than some of the teachers, who may thus become insecure, evasive, or even antagonistic. When teaching cancer epidemiology, it is usually required that students have had at least an introductory course of epidemiologic principles and methods. In practice, however, understanding of basic epidemiologic concepts may require more than an introductory course, and students will usually welcome the presentation of methodological issues as applied to specific aspects of cancer epidemiology.

The invocation of a coherent picture, even as a working hypothesis, is highly conducive to a successful presentation. There is nothing more unpleasant for students and teachers alike than to face a list of fragmentary data that cannot be integrated, even tentatively. For several forms of cancer, however, the information is so limited (e.g. brain cancer) or so conflicting (e.g. lymphomas) that it becomes impossible for a teacher to present a coherent picture on the basis of the available information. Nevertheless, every effort should be made to provide such a coherent picture, always making sure that the students are aware of the feebleness of the supporting evidence.

Goal

The goal of a fifteen-to-twenty-hour course in cancer epidemiology is to integrate simple principles of biology with epidemiologic characteristics and concepts so that the discipline moves beyond biostatistics into the realm of

biomedical sciences. Indeed, it has been argued that the relation of epidemiology to biostatistics is similar to that between physics and mathematics. In other words, the goal of cancer epidemiology is not only to describe the principal risk factors of various forms of cancer but to evaluate the compatibility of the risk profile of a particular cancer with alternative biologic hypotheses.

Course structure

In this section, we present an option for the structure of the course, which obviously can be adjusted to accommodate the experience, expertise, and style of the instructor, as well as the background, interest, and pace of the class.

Session 1: review of biology

It is usually useful to start a course on cancer epidemiology by reviewing principles of cancer biology. This is in most situations a very difficult session because some of the students may have a more solid understanding of biology than their teacher, while others will not be able to get an adequate grasp of the issues within two or three hours. It is advisable to request that, for the first session, students must have read a simple overview, such as the chapters on the origin of cancer (Ecsedy and Hunter 2008) and the genetic epidemiology of cancer (Haiman and Hunter 2008) from *Textbook of Cancer Epidemiology* (Adami et al. 2008), or chapters of similar content from several other excellent textbooks. Similar review articles frequently appear in journals addressed to educated but not specialized readership.

During the first and relatively long session, several issues need to be touched upon, because they need to be invoked during the teaching of cancer epidemiology. These are the monoclonal nature of most cancers; the importance of apoptosis in carcinogenesis; antigenic and biochemical differences between normal and cancer cells; in vitro transformation; the cell cycle and control points; genetic and epigenetic phenomena in carcinogenesis; stages of carcinogenesis, and experimental and epidemiologic evidence that cancer is a multistage process; chemical carcinogenesis; precarcinogens and ultimate chemical carcinogens; DNA repair deficiency; nature and mode of action of oncogenic viruses; oncogenes; chromosomal aberrations, including constitutional changes (e.g. retinoblastoma); heritability of cancer predisposition and chromosomal rearrangements; polymorphic systems and their potential importance in environmental carcinogenesis; immunologic phenomena in carcinogenesis; short-term in vitro tests for carcinogen detection; rodent and other animal bioassays for carcinogen detection; and limitations in the prediction of human carcinogens on the basis of in vitro tests and animal bioassays.

Biologic plausibility is essential for the interpretation of epidemiologic data. Two errors can be made in this context. The first is simply to ignore biology, while the second is to argue that something is biologically plausible simply because it is conceivable. For instance, the nature of prions was implausible but not inconceivable. In contrast, some epidemiologic theories should have been labelled as inconceivable as soon as it was recognized that they tend to contradict simple laws of physics or chemistry.

Session 2: causality in cancer epidemiology

The second session could be devoted to the use of epidemiologic methods in the documentation of cancer etiology. There are many excellent introductory and intermediate-level epidemiology books available. Comprehensive reviews with a special focus on cancer are also available from several authors (e.g. Lagiou et al. 2005; Lagiou, Trichopoulos, et al. 2007).

Because of the rising interest in the genetic epidemiology of cancer, a brief introduction into studies exploring the genetic origins of cancer is necessary. Here, the challenge is to provide a simple overview and to avoid confusing the students with the complicated terminology that permeates the genetic epidemiology literature. The outline indicated below relies on a relevant section by Lagiou, Trichopoulos, et al. (2007) from *The Cancer Handbook* (Alison 2007).

Two main types of epidemiologic studies are used for the identification of genes predisposing to cancer: genetic linkage studies and genetic association studies. Genetic linkage studies are generally undertaken in families with several cancer cases and rely on the principle that two genetic loci, or a cancer and a marker locus, are linked when they tend to be transmitted together from parent to offspring more often than expected by chance. Linkage extends over large regions of the genome and refers to a locus, rather than specific alleles in that locus, which can vary from family to family (Teare and Barrett 2005). Such studies have led to the identification of genes that have substantial impact on the occurrence of breast cancer and colorectal cancer, although such genes are generally rare.

Genetic association studies can be of either cohort or case-control design. They are frequently undertaken in the general population and are conceptually similar to traditional epidemiological investigations. The difference is, however, that instead of focusing on environmental factors, genetic association studies evaluate as 'exposures' specific alleles (rather than loci) of genetic polymorphisms, usually single nucleotide polymorphisms (SNPs). The specific alleles may be etiologically related to cancer or, much more frequently, very closely linked to the truly etiological allele, which may not be known. The actually

investigated allele and the true etiological allele are said to be in linkage disequilibrium—that is, they are so closely linked that they tend to be inherited together. Linkage disequilibrium-based genetic association investigations cover shorter genetic regions than linkage studies (Cordell and Clayton 2005; Teare and Barrett 2005). Occasionally, however, the specific allele may be chosen for study because it is thought to be involved in the etiology of the cancer under investigation (e.g. a candidate gene). Many SNPs over large parts of the genome, or even over the whole genome, may also be evaluated, with little or no prior evidence that most of them are etiologically relevant or are in linkage disequilibrium with etiologically relevant genes. Genetic association studies have not been very successful to date in identifying genes or polymorphisms involved in cancer etiology, possibly because the respective relative risks deviate very little from the null value but also because the tools to examine alleles across the genome simultaneously are only just becoming available (Wacholder et al. 2004; Hunter 2012).

Session 3: cancer incidence around the world

The third session should be devoted to an overview of cancer incidence around the world. Data on cancer incidence from cancer registries have been regularly published by the International Agency for Research on Cancer (IARC) in the volumes *Cancer Incidence in Five Continents* (Curado et al. 2007) and are also available online (Ferlay et al. 2010). A good approach is to present data from a few, long-standing, well-functioning registries and to focus on major sites of cancer, so as not to overwhelm the students with heavy and difficult-to-follow tables. Such registries are available for Europe (Denmark (national) and Sweden (national)), North America (USA Surveillance Epidemiology and End Results), and Asia (Osaka in Japan). The indicated registries cover, to a considerable degree, the existing variation for cancers at different sites throughout the world.

It appears preferable to present cumulative incidence rates that can be interpreted as probabilities and are usually expressed in percentages. The cumulative incidence rates cover the age span 0–74 years and are easier to understand than standardized incidence rates. Thus, many people are able to determine that the cumulative incidence of breast cancer among Caucasian women in the US is more than 10 per cent but is less than 5 per cent among women in China or Japan. In contrast, few investigators, let alone students, are able to provide figures for the age-adjusted incidence of breast cancer in the US and China or Japan. The students may be asked to attempt to explain the observed differences, and the resulting discussion is generally lively and occasionally quite insightful.

Session 4: overview of cancer causation

The fourth session could cover the existing evidence on cancer causation. The teacher may first wish to present a table (Table 17.1) of some relatively common occupational agents that have been characterized by IARC as human carcinogens. It should be pointed out to students that many more occupational exposures have been characterized as established or probable human carcinogens but that these exposures are now well controlled in developed countries so that no more than 5 per cent of all cancer deaths are attributed to them. It is instructive to point out to the students that the success of epidemiology in identifying many occupational carcinogens is due to tragic natural experiments—exposure to very high concentrations of carcinogens in occupational settings in the past.

Table 17.1 Important occupational carcinogens.

Occupational exposure	Cancer site	General population exposure	Main professions exposed
Arsenic	Lung, skin	Uncommon	Insecticide and herbicide sprayers; tanners
Asbestos	Mesothelioma, lung	Uncommon	Brake-lining, insulation, and demolition workers
Benzene	Myelogenous leukaemia	Common	Painters; distillers and petrochemical workers; rubber workers
Diesel exhaust	Lung	Common	Railroad and bus-garage workers; truck operators
Man-made mineral fibres	Lung	Uncommon	Wall and pipe insulation
Hair dyes	Bladder	Uncommon	Hairdressers and barbers (inadequate evidence for customers)
Mineral oils	Skin	Common	Metal machining
Non-arsenical pesticides	Lung	Uncommon	Sprayers; agricultural workers
Painting materials	Lung	Uncommon	Professional painters
Soot	Skin	Uncommon	Bricklayers; insulators; firefighters; heating-unit workers

Adapted with permission from D. Trichopoulos et al., 'What causes cancer?' *Scientific American*, p.84, Copyright © 1996 Scientific American, a Division of Nature America, Inc.

Table 17.2 Biological factors considered as carcinogenic in humans by the International Agency for Research on Cancer.

Factor	Cancer site
Aspergillus flavus (aflatoxins)	Liver
Hepatitis B virus	Liver
Hepatitis C virus	Liver
Helicobacter pylori	Stomach
Schistosoma hematobium	Urinary bladder
Opisthorchis viverrini	Cholangiocarcinoma
Human papilloma viruses (HPV)	Cancer of cervix and anus
Herpes simplex virus 8 (HSV-8)	Kaposi's sarcoma
HIV-1	Non-Hodgkin's lymphoma, Kaposi's sarcoma, anus
HTLV-1	Adult T-cell leukaemia
Epstein–Barr	Lymphoproliferative diseases, nasopharynx

Table 17.2 presents the biologic agents characterized by IARC and the wider scientific community as carcinogenic. The teacher could indicate that infectious agents may account for up to 20 per cent of all cancer deaths in the developing countries, although much less in the developed ones. *Aspergillus flavus* is included in this category because it contaminates stored foods and produces the liver carcinogen aflatoxin.

Table 17.3 shows lifestyle factors, excluding nutrition, that have been established as human carcinogens. These factors account for about 25 per cent of all cancer deaths around the world; most of these deaths are due to tobacco smoking. Table 17.4 refers to the role of diet and related factors in the etiology of various forms of cancer (World Cancer Research Fund and the American Institute for Cancer Research 2007). There are no formal IARC evaluations for nutritional factors. Although case-control studies have provided evidence for an inverse association of vegetables and/or fruits with several forms of cancer, cohort studies have indicated considerably weaker associations (Hung et al. 2004; Potter 2005; Boffetta et al. 2010). At this point, the teacher could comment that it is not clear whether case-control studies suffer collectively from selection and/or information bias, or whether non-differential exposure misclassification in cohort studies, which usually rely on self-administered questionnaires, drives existing associations towards the null. Nevertheless, it appears that diet in adult life (as well as consequent body weight) is responsible for

Table 17.3 Lifestyle factors considered as carcinogenic in humans by the International Agency for Research on Cancer.

Factor	Cancer site
Smoking	Lung, urinary bladder, oesophagus, oral cavity, larynx, kidney, pancreas, liver, other
Passive smoking	Lung
Ethanol	Oesophagus, larynx, oral cavity, pharynx, liver, breast, large bowel
Betel chewing	Oral cavity
Ionizing radiation	Bone marrow and several other sites
Ultraviolet radiation	Skin, lip

Reproduced with permission from P. Lagiou et al., Causality in cancer epidemiology, *European Journal of Epidemiology*, Volume 20, Issue 7, pp. 565–74, Copyright © 2005 Springer.

Table 17.4 Risk implications for major forms of cancer by diet-related factors.*

Cancer site	Vegetables/ Fruits	Meat (red)	Alcohol	Salt	Hot drinks	Height	Obesity	Physical activity
Mouth/pharynx	↓↓		↑↑↑					
Nasopharynx	↓			↑↑				
Oesophagus	↓↓	↑	↑↑↑		↑		↑↑↑	
Stomach	↓↓		↑	↑↑				
Large bowel	↓	↑↑↑	↑↑↑			↑↑↑	↑↑↑	↓↓↓
Liver			↑↑↑				↑	
Gallbladder							↑↑	
Pancreas	↓	↑				↑	↑↑	↓
Larynx	↓↓		↑↑↑					
Lung	↓↓						↑	↓
Breast	↓		↑↑↑			↑↑↑	↓↓/↑↑↑†	↓↓
Endometrium	↓					↑	↑↑↑	↓
Ovary						↑		
Prostate	↓	↑				↑	↑	↓
Kidney							↑↑	

* ↓↓↓ inverse association, ↓↓ probable inverse association, ↓ possible inverse association, ↑↑↑ positive association, ↑↑ probable positive association, ↑ possible positive association.

† Inverse among premenopausal women; positive among postmenopausal women.

Data from P. Lagiou et al., Causality in cancer epidemiology, *European Journal of Epidemiology*, Volume 20, Issue 7, pp. 565–74, Copyright © 2005 Springer and P. Lagiou et al., Nutritional epidemiology of cancer: Accomplishments and prospects, *Proceedings of the Nutrition Society*, Volume 61, Issue 2, pp. 217–222, Copyright © The Authors 2002.

about 15 per cent of all cancer deaths worldwide. This estimate does not include the consequences of excess energy intake and other factors in early life (Trichopoulos et al. 1996; Ahlgren at al. 2004; Lagiou, Adami, et al. 2007).

Medical products and procedures are responsible for about 2 per cent of cancer deaths worldwide but it is generally accepted that the effectiveness of these procedures outweighs their cancer risks. Three major categories of pharmaceutical agents are considered as increasing cancer risk: cancer therapeutic drugs, immunosuppressive drugs, and exogenous hormones. Among the iatrogenic processes that increase cancer risk, one also needs to consider radiotherapy. With respect to exogenous hormones, it should be taken into account that they may increase the risk for some cancers while decreasing the risk for others. For instance, oral contraceptives reduce the risk for ovarian cancer but may slightly increase the risk for breast cancer.

The teacher must stress that there are endogenous factors that affect cancer risk; some of these factors are already known but others are unknown, and many are uncontrollable. Reproductive factors are of critical importance for breast, ovarian, and endometrial cancers, whereas sexual activity is an important risk factor for cancer of the cervix. Perinatal factors, and particularly early growth, may affect the risk of several forms of cancer, perhaps by modulating the number of susceptible stem cells. Finally, two categories of genes need to be considered: major genes, frequently dominant with high penetrance, like *BRCA1* and *BRCA2*, which may be responsible for less than 5 per cent of all cancer deaths; and genetic polymorphisms that are likely to interact with a wide range of environmental factors in the causation of an unknown but probably more substantial proportion of cancer cases and deaths worldwide. Students should appreciate that natural selection tends to limit the prevalence of powerful cancer-causing genes much more than it affects the prevalence of interacting alleles in polymorphic systems.

Table 17.5 shows the estimated proportions of worldwide cancer deaths that may be attributed to environmental factors (data from the 2000s). Students should be aware that attributable fractions are, in general, non-additive because several etiologic factors act synergistically. Some teachers may wish to conclude the session by pointing out the dominant role epidemiology has played in the identification of most causes of cancer and by also discussing the fact that, for a fraction of all cancers, the causes may never be identified or may turn out to be unavoidable, possibly because they are intimately linked to natural growth phenomena.

Session 5: liver cancer

No other cancer epidemiology is as intriguing to students as that of liver cancer. Although rare in developed countries, liver cancer is very common in

Table 17.5 Maximal potential reduction of cancer mortality in the world by categories of factors.

Factor or category of factors	Percentage
Tobacco	25
Alcohol	3
Diet in adult life (including obesity)	15
Excess energy intake in early life	5
Food additives and contaminants	1
Sedentary life	2
Infectious agents	20
Reproductive factors	4
Ionizing and ultraviolet radiation	2
Occupational factors	5
Environmental pollution	3
Medical products and procedures	2

sub-Saharan Africa and South East Asia, areas of the world which young, motivated students are interested in and care about. It is also the first major human cancer to be identified as having a dominant viral etiology. Preventive medicine is of particular importance for liver cancer, because this cancer is frequently incurable but largely preventable with current knowledge and technology. As far as teaching is concerned, the discovery of the causes of liver cancer is an ideal example for the demonstration of use of epidemiologic methodology for the identification of cancer causes. The discovery of the etiologic role of hepatitis B virus (HBV), in particular, has followed an epidemiological textbook sequence, from case reports and geographical observations to case-control studies and cohort investigations.

Students should learn that primary liver cancer is distinguished into three histological types: the by far most common hepatocellular carcinoma, the rare cholangiocarcinoma, and the very rare angiosarcoma. The distinction is crucial because the three histological types have distinct etiologies. For hepatocellular carcinoma, the established causes are HBV, the hepatitis C virus (HCV), aflatoxins, alcoholic and other forms of cirrhosis and, convincingly, although not widely appreciated, tobacco smoking. For cholangiocarcinoma, the recognized causes are *Opisthorchis viverrini* and perhaps *Clonorchis sinensis*, both of which are flukes prevalent in South East Asia. Last, for angiosarcoma, thorotrast, vinyl

chloride monomer, inorganic arsenic, and androgenic anabolic steroids are considered to be causes.

The teacher should structure the presentation on the basis of a particular text that reflects his/her own style and expertise. The chapter by Stuver and Trichopoulos (2008) on liver cancer is a useful source.

Session 6: lung cancer

Lung cancer in relation to smoking has been traditionally used as a teaching example in epidemiology. Most students, however, consider the topic somewhat boring. The teacher may reinstate their interest if the problem is presented as follows: at least half and perhaps two-thirds of the lung cancer cases around the world are caused by smoking. However, until the smoking epidemic is controlled, it would be of interest to know *who* among the smokers are likely to develop lung cancer and *why* some of the non-smokers end up developing the disease. It is also a surprise for many students to realize that lung cancer among non-smokers is a problem of comparable magnitude to that of cancer of the pancreas or ovary.

Introduction of the descriptive epidemiology of lung cancer is facilitated by presenting the figures regularly produced by the American Cancer Society and published every year in the January issue of *CA: A Cancer Journal for Clinicians* (e.g. Siegel et al. 2013). These figures effectively convey the magnitude of the lung cancer problem in developed countries, in which more than a quarter of cancer deaths among men and almost one-fifth of cancer deaths among women are due to lung cancer (American Cancer Society 2011).

The evidence linking tobacco smoking with lung cancer can be a fascinating story, going through population correlations to case-control investigations and major cohort studies. Invocation of experimental evidence and molecular discoveries can further enrich the presentation. Although everybody knows that smoking causes lung cancer, many aspects of this relation and the methodological twists that were required for their identification are rich teaching material. Among the issues to be addressed are the comparative lung carcinogenicity of cigarettes, pipes and, cigars; the modifying effects of inhalation; the variable lung cancer risk in relation to number of cigarettes smoked and mouth-keeping time; the substantially higher risk among early starters of this habit; the disproportionally high effect of duration of smoking on lung cancer risk; and the consequences of the introduction of low tar and filter cigarettes, as well as e-cigarettes. This session can be concluded by pointing out the success, however limited, of smoking cessation. For more advanced classes, the teacher can indicate to the students how the exposure–response patterns between smoking and lung cancer risk have contributed to

the estimation of the tobacco initiation and tobacco cessation lung cancer latencies; the multiple stages of the lung cancer carcinogenesis; the absence of a threshold in the exposure–response curve; and other subtle and intriguing issues.

Although invocation of smoking is a reflex in lung cancer epidemiology, the teacher should also point out that many occupational carcinogens have the lung as their target. Among them, asbestos, radon, and polycyclic aromatic hydrocarbons are of major and continuing interest. In the context of ionizing radiation effects, it is useful to present the comparative risks of alpha particles, beta particles, and gamma radiation as a function of their relative biologic effectiveness. Although occupational causes of lung cancer have been well documented in the context of the conditions prevailing in factories of developed countries several years ago, they collectively explain now a small and declining proportion of lung cancer cases. In contrast, the effect of a factor of potential importance for a non-negligible fraction of lung cancer cases has remained undocumented. This is air pollution, the prolonged exposure to which is very difficult to measure accurately. The conflicting evidence from ecologic and analytical epidemiologic studies, together with the biologic plausibility of an effect of air pollution on lung cancer risk, creates a rich but complex field for methodological explorations and biologic considerations.

There is evidence that intake of fruit and vegetables may reduce the risk for lung cancer but attempts to attribute this effect to vitamin A and beta-carotene have failed. Randomized trials have not supported a beneficial effect of micronutrients hypothesized to protect from lung cancer on either epidemiologic or biomedical evidence. This raises the issue of limitations in extrapolating from foods to nutrients.

The relation of passive smoking to lung cancer risk also needs to be considered. The epidemiologic evidence linking passive smoking to lung cancer risk points to a weak positive association. However, because smoking is an established carcinogen and there is no threshold in the exposure–response association between smoking and lung cancer risk, it is reasonable to assume that passive smoking affects lung cancer risk. Whether the excess risk for lung cancer among individuals exposed to environmental tobacco smoke (ETS) is 30 per cent or higher than that in individuals not exposed to ETS has been the subject of an interesting controversy involving both epidemiologic evidence and biomedical plausibility.

The chapter on lung cancer by Boffetta and Trichopoulos (2008) is a useful source of information for a prospective teacher but there are many other good reviews of lung cancer epidemiology and etiology (Schottenfeld and Fraumeni 2006).

Session 7: breast cancer

After the presentation in distinct sessions of two success stories of epidemiology with respect to liver and lung cancer, the teacher should share with students the reality of failure, all too frequent in real life. Indeed, given the overall stability or even increase of breast cancer incidence rates over time around the world, it would be difficult to disagree with the statement that, although abundantly funded, breast cancer research has not produced sufficiently useful results for the primary prevention of the disease. Because the overall situation is at present so fluid, teachers generally follow the approach they feel more comfortable with. The approach presented in this section reflects, to a certain extent, the context in which the authors of this chapter are currently working, which may or may not be similar to the context in which other investigators are working.

Established breast cancer risk factors

The epidemiology of breast cancer has been reviewed by many investigators, and the following risk factors are generally recognized (Hankinson et al. 2008):

- breast cancer is about 100 times more common among women than among men;
- the incidence of breast cancer increases with age, with a characteristic inflection around the age of menopause;
- early age at menarche is associated with increased breast cancer risk;
- early age at menopause is associated with reduced breast cancer risk;
- for a given age, a surgically induced menopause conveys more protection than a naturally occurring one;
- the earlier the age of the first full-term pregnancy, the lower the risk for breast cancer (subsequent full-term pregnancies have a similar but quantitatively much weaker effect);
- after the age of approximately 35 years, the occurrence of a first pregnancy increases rather than reduces breast cancer risk;
- prolonged lactation conveys some protection against breast cancer risk over and beyond that conveyed by increased parity (although the effect is fairly small);
- height is positively associated with breast cancer risk;
- obesity is inversely related to breast cancer risk among premenopausal women but positively related among postmenopausal women;
- the risk of breast cancer is threefold higher for Caucasian women in Europe and North America than for women in eastern Asia;

- the risk indicated by a high-density mammogram is almost fourfold higher than that indicated by a low-density mammogram;
- breast cancer tends to be slightly more common among women of higher socio-economic status and among urban rather than rural residents;
- major genes that convey increased susceptibility to breast cancer have been identified but such genes are responsible for less than 10 per cent of breast cancer cases overall;
- current or recent use of oral contraceptives slightly increases the risk for breast cancer;
- current or recent use of long-term combined hormone-replacement therapy increases the risk for breast cancer;
- ionizing radiation is an established cause of breast cancer but of limited quantitative importance;
- consumption of alcoholic beverages slightly increases breast cancer risk;
- adult life diet and physical activity have some, albeit limited, effect on breast cancer risk; and
- there is no compelling evidence that exposure to organochlorines or electromagnetic fields affects breast cancer risk.

An etiologic model

A synthesis of existing evidence about breast cancer has led to the formulation of a hypothesis with three key components (Trichopoulos et al. 2008):

1. the likelihood of breast cancer occurrence depends on the number of mammary tissue-specific stem cells, a number which is determined early in life, notably in utero or in the immediate postnatal life;
2. in adult life, all growth-enhancing mammotropic hormones, in conjunction with their receptors, affect the likelihood of retention of cells with spontaneous somatic mutations, as well as the rate of expansion of initiated clones; and
3. while a pregnancy stimulates the replication of already initiated cells, it conveys long-term protection through the terminal differentiation of a large fraction of the mammary tissue-specific stem cells.

The etiologic model outlined above does not contradict any of the hypotheses that are currently dominant and is a synthesis rather than an original construct. Its advantage in a teaching process is that it provides biologic underpinnings for the whole spectrum of empirical evidence and specifies targets for refutation, in line with the mainstream Popperian approach.

Session 8: diet and cancer

Cancer epidemiology has been generally successful, from the discovery of tobacco smoking and HBV as the most important causes of two common forms of cancer, to the identification of asbestos and several other carcinogens in various occupational settings. This has created optimism in the scientific community and the general public about the prospects of clarifying the role of nutrition in the etiology of several cancers. However, in these successful efforts, epidemiologists were trying to assess the effects of single, easily identifiable factors. Furthermore, these effects were large in the relative sense, either inherently (HBV) or because the exposure levels in particular settings were extreme (asbestos in insulation workers). In contrast, it is difficult to assess nutritional factors over a long period of time, it is difficult to determine the relative importance of food items, food groups, particular nutrients, and nutritional patterns, and the relative risk associated with any particular nutritional factor is likely to be small (extreme contrasts can be found only with respect to supplemental micronutrients). Nevertheless, for several common cancers, the large inter-population variation or evidence from migration studies cannot be attributed to genetic factors or accounted for by other established environmental factors or personal attitudes.

In the session dedicated to the nutritional epidemiology of cancer, the emphasis should be on the methodological complexities of epidemiologic studies in this field. The validity of long-term diet ascertainment is known to be limited. This has long been recognized but it has usually been assumed that it would generally lead to findings indicating that there are no associations, when actually there may be real associations, albeit small ones (false-negative results). However, given the extensive intercorrelation between food items and food groups, extensive mutual confounding should also be expected, and residual confounding can cause both false-negative and false-positive results. Furthermore, even without confounding, extensive misclassification can reduce the statistical power to such an extent that the predictive power of a particular 'positive' finding will be low. The situation can be further complicated when the latencies of particular nutritional factors are different or operationally non-definable, or when thresholds must be invoked.

Another complex issue focuses on the components of variability of particular nutritional intakes (intra-individual, inter-individual, and time-dependent variability). Important work has been undertaken in this field (Willett 2013) but many problems remain unresolved. Additional complexity is introduced by the dependence of the relative risk associated with a particular nutritional factor on the presence of other nutritional component causes in the same or in different etiologic complexes.

Current knowledge on the nutritional etiology of human cancer has been carefully compiled and summarized in a large volume that has been prepared by the World Cancer Research Fund and the American Institute for Cancer Research (2007). This is a reference book but it also provides summary tables for nutritional factors in the etiology of various forms of cancer.

Session 9: a flexible session

The seventh session may be a flexible one and depends on the background and interest of the students. If several students have had difficulty in following the course, the session could be dedicated to reviewing the material covered, enriched with additional methodological issues or substantive information.

For students who are interested in methodology or have a strong background in physics, it might be useful to consider the evidence concerning the undocumented carcinogenicity of extremely low frequency magnetic fields and radiofrequencies. In this area the empirical evidence is conflicting but the methodological issues are of particular interest because they emerge from efforts to distinguish between null effects distorted by chance, confounding, and bias, and weak effects attenuated by poor exposure ascertainment and latency mis-specification.

For students with a strong background in biology, it might be preferable to discuss the epidemiology of colorectal cancer, because several epidemiologic findings are well documented and there is also an adequate understanding of the molecular events that underlie the natural history of this cancer (Vogelstein and Kinzler 2004). A useful reference is the relevant chapter by Potter and Hunter (2008) or the corresponding chapter in the book edited by Schottenfeld and Fraumeni (2006).

Another alternative is to discuss the epidemiology of prostate cancer, a tumour about which little is known and for which study design issues are complicated by the expanding use of screening for prostate-specific antigen (PSA) and by the large fraction of cases in which the histologically indistinguishable disease tends to remain latent and subclinical.

Session 10: occupational cancer

The epidemiology of occupational cancer is a topic rich in substantive knowledge. Indeed, most established causes of cancer have been identified through epidemiologic studies done in settings where extremely high exposures have generated very high excess risks for specific forms of cancer. It is again advisable to devote most of the available time to methodological issues and discuss only briefly the principal occupational carcinogens and their associated cancers, as indicated, for example, in Table 17.1.

Since the documentation of particular agents, mixtures, or processes as carcinogenic has mostly relied on tragic 'natural experiments', it is useful to compare genuine experimental studies in laboratory animals with observational studies involving humans and being harvested by occupational epidemiologists. The advantages of animal experiments are the detailed definition of exposure and of possibly interacting factors; the careful establishment of the outcome; the free use of randomization, with fewer methodological concerns and less strict ethical constraints; and the exploitation of animal inbreeding that tends to produce animals that are genetically susceptible, as well as of a similar genetic background, and thus have a smaller response variability than human populations. Against these advantages, animal experiments also have considerable disadvantages. Because the animal populations cannot be very large, exposures have to be unrealistically high, a condition which further complicates the already substantial biologic barriers for interspecies extrapolations.

Most studies of occupational epidemiology are by definition relevant to humans, except when working conditions have changed so rapidly and so substantially as to make the results of early occupational epidemiologic studies irrelevant to current conditions, mainly because of different orders of magnitude in exposures. However, occupational epidemiologic studies are not as straightforward as they are generally thought to be for various reasons:

- occupations appear and disappear, work tasks change and are often not comparable over time or between countries, and there may be questionable correspondence between occupational titles and actual exposures;
- it is difficult to know whether short-term peaks or time-weighted averages are more important;
- there may be limited comparability between exposed and non-exposed individuals, particularly when an outside population is used for the calculation of expected cases;
- outcome data may be of questionable quality or limited to death certificates;
- information on confounders, except gender and age, is usually missing; and
- actual measurements of the exposures under consideration are rarely available, so duration of exposure is used as a proxy for exposure quantification—in this instance, time is used in at least two different ways: one for estimating cumulative exposure level, and another for estimating the total person-time at risk.

Given the complexity of these issues and the considerable obstacles in undertaking a sound epidemiologic study, it may be surprising that occupational

epidemiologic studies have been so successful. This is for two reasons: (1) the effect estimate (e.g. relative risk) is so high as to override any concern about confounding and bias, and (2) the carcinogenic effect is so specific as to leave very little doubt about the underlying validity (e.g. the causation of mesothelioma after asbestos exposure is very difficult to explain in terms of confounding and bias).

Several other methodological issues need to be addressed even in a medium-level course. The student should understand that retrospective cohort studies with internal or external comparison groups are, in most instances, the only realistic option for the calculation of relative risk. Students should also be warned that confounders, beyond age and gender, can rarely generate confounding rate ratios of more than 1.5.

Occupational epidemiology provides the best context for learning the applicability and the interpretability of exposed attributable risks (or rates) and for understanding the legal implications of an exposed attributable risk of higher than 2.0 (it is more likely than not that a particular exposure was the cause of the relevant outcome in an exposed person, given the baseline risk of 1.0). The student should also recognize that, although risk ratios and exposed attributable risks (or rates) can be estimated in retrospective cohorts, this design can rarely generate population-attributable risks (or rates), because the latter effect parameter requires an unbiased estimate of the exposure prevalence.

The issue of the proportion of cancer cases in a population that can be attributed to occupational exposure has been estimated in developed countries to be around 5 per cent. It may be smaller, however, because conditions in the workplace improve over time in most countries. It may also be higher if occupational carcinogens have not been identified because the hazard has been unsuspected, the relative risk (or rate ratio) is small, there are few exposed individuals in relevant working places, and/or the latency of the particular cancer is long. Little is known about occupational cancers in developing countries.

The epidemiology of occupational carcinogenesis is a complex topic, mainly because time is considered in several dimensions but also because the quality of data is rarely satisfactory. The students should realize that relative risks below 1.5, when generated from inherently poor studies, should be interpreted with caution. They should also realize that, whenever they accept an unsubstantiated elevated risk at face value because workers should have the benefit of the doubt, they may actually harm them by failing to identify the true cause of the excess risk.

In addition to several excellent textbooks, useful reviews have also been published on this topic (Boffetta 2004; Siemiatycki et al. 2004).

Session 11: screening for cancer

Whatever the duration and the content of a cancer epidemiology course, it is proper to conclude it with the principles of screening. Screening is an unusual process in that it philosophically belongs to prevention and methodologically relies on epidemiology but cannot be implemented without some form of effective intervention which has the hallmarks of therapeutic activity. Screening is in the epicentre of all secondary prevention activities, although, when screening is focused on risk factors, it may also be thought of as a primary prevention process. Screening activity has been defined as the presumptive identification, with simple means, of those among healthy people who are likely to be diseased. The screening activity should not be confused with the screening test, which is simply a test that is applied without clinical indication. Screening activity should also be distinguished from the timely recognition of cancer symptoms or signs, which is a manifestation of good clinical practice and has as an objective the recognition of cancer at an earlier but already clinical stage.

The philosophy of screening can be operationalized by invoking the concept of the critical point. The critical point is that point in the natural history of cancer beyond which the disease becomes essentially uncontrollable; for example, because metastases have taken root. If the critical point is beyond the time of usual clinical diagnosis, there is no reason for screening because the disease can be effectively treated even after the appearance of clinical disease, for example, non-melanoma skin cancer. Screening is also pointless when the critical point occurs before the screening-generated diagnosis, as, for example, with respect to hepatocellular carcinoma. In contrast, screening is valuable when clinical diagnosis occurs after the critical point, whereas screening diagnosis may take place before that point. The latter situation exists for screening of colorectal, cervical, and breast cancer. Indeed, in about 20 per cent of breast cancer cases, the critical point appears to lie between the point of clinical diagnosis and the point of screening diagnosis, so that one-fifth of women may be effectively treated with proper application of screening procedures.

Application of screening requires some conditions that are linked to the disease under consideration, the available screening test, and the overall organization of the screening programme. The disease under consideration should be frequent, sufficiently serious, and have a detectable preclinical phase during which the critical point is frequently located. The screening test has to be simple, quick, inexpensive, safe, and acceptable to the population and it should be characterized by high repeatability as well as high validity, the latter requiring high values of both sensitivity and specificity. Finally, a crucial programme condition is that the detectable preclinical phase of the disease under consideration

should have a relatively high prevalence. This is because, for given values of sensitivity and specificity of a test, the prevalence of the detectable preclinical stage is a crucial determinant of the all-important positive predictive value of the test under the programme conditions. Programme conditions also include the compliance of the population, the availability of resources, and a competitive cost–effectiveness ratio.

For the evaluation of a screening programme, there are several necessary but not sufficient criteria, including the yield of preclinical cases and prognostic indicators, such as the stage distribution and the fatality of the screening-detected disease. Shifting the stage distribution to the left and reducing the fatality ratio of screening-detected cancer cases are necessary but cannot in themselves document the effectiveness of a screening programme. To establish this, the mortality rate from the disease under consideration needs to be reduced, because neither lead time bias nor length time bias, the two biases that compromise the prognostic indicators, can possibly affect mortality.

Only mortality rate is a valid outcome, but a valid design is also needed to generate a sound outcome. This requires randomization of the total population, because non-random distribution carries the risk of bias generated by the tendency of screened volunteers to have a favourable cancer stage distribution, quick response to disease warnings, and good compliance.

Most students already have some knowledge of the principles underlying screening for cancer, so that the session on screening is, in terms of topic complexity, a decrescendo. It should be made clear to students, however, that screening is one of the most complex medico-social undertakings and that those who want to specialize in the epidemiologic aspects of screening should be ready for major intellectual challenges.

For an introductory session of cancer screening, one may wish to look at some early writings (Morrison 1992) and at the relevant chapters of the textbook by Schottenfeld and Fraumeni (2006).

Assessing students' achievements

A series of multiple-choice questions distributed to the students for self-evaluation after, say, the fourth, sixth, and eighth sessions and subsequently discussed in class serves three purposes:

- it allows the instructor to assess and adjust the pace of the course;
- it allows students to evaluate their progress; and
- by repeating the material covered, it refreshes knowledge and facilitates understanding.

A subsample of the multiple-choice questions, perhaps ten questions out of a total of between sixty and one hundred, may be given to the students during the final examination, which may be oral or written.

Have the objectives of the course been met?

At the end of the course the students should be able to answer questions like the following:

- What are the principal stages in the natural history of neoplasias? (Initiation by genotoxic agents, promotion by conditions that favour cellular multiplication, progression under the influence of factors frequently involved in normal processes.)
- Which type of oncogenic virus is intimately linked to oncogenes? (RNA viruses.)
- Which is the common characteristic of infectious agents that are carcinogenic to humans? (They cause chronic infection.)
- Which characteristic makes a form of radiation definitely carcinogenic? (Ionization.)
- Which forms of cancer have clear long-term trends? (Stomach cancer is decreasing in both genders, melanoma is increasing in both genders, etc.)
- Why is a causal factor ranked differently depending on the use, as a criterion, of the relative or the population-attributable risk? (Prevalence of exposure matters only for attributable risk.)
- Why is the incidence of lung cancer among women in some populations lower among people older than seventy years than among younger people? (Cohort or generation effect.)
- Do sophisticated statistical techniques always remove confounding? (No, because confounders may be unknown or poorly measured.)
- Can the very high mortality rate from AIDS in some African countries reduce age-adjusted cancer mortality rates in these countries? (No, because, as a rule, rates do not compete.)
- Can the very high mortality rate from AIDS in some African countries reduce the risk for dying from cancer in these countries? (Yes, because risks do compete.)
- Can magnetic fields cause cancer as initiators or as growth enhancers? (Impossible as initiators, conceivable but implausible as growth enhancers.)
- Which epidemiologic design is usually more appropriate for the documentation of the carcinogenicity of a particular occupational exposure?

(Retrospective cohort study of a special exposure group with an internal reference group.)

- Which prognostic indicator is appropriate for the evaluation of a randomly allocated screening programme? (None, because of lead time bias and length time bias.)
- What are the most important achievements in cancer epidemiology? (Smoking and lung cancer, hepatitis viruses and liver cancer, occupational carcinogens, etc.)

Cancer epidemiology is a relatively young discipline that was mostly developed after the Second World War. Although, it has made major contributions, the remaining challenges are enormous. Teaching and studying this discipline are particularly rewarding because, in addition to sharing and advancing knowledge on the specific topics, such activities contribute to the understanding and development of epidemiologic methodology.

References

Adami, H. O., Hunter, D., and Trichopoulos, D., eds. (2008) *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. xxxiii + 748.

Ahlgren, M., Melbye, M., Wohlfahrt, J., and Sørensen, T. I. (2004) Growth patterns and the risk of breast cancer in women. *New England Journal of Medicine*, **351**: 1619–26.

Alison, A. R., ed.-in-chief. (2007) *The Cancer Handbook* (2nd edn). Chichester: John Wiley & Sons.

American Cancer Society. (2011) *Global Cancer Facts and Figures* (2nd edn). Atlanta: American Cancer Society.

Boffetta, P. (2004) Epidemiology of environmental and occupational cancer. *Oncogene*, **23**: 6392–403.

Boffetta, P. et al. (2010) Fruit and vegetable intake and overall cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Journal of the National Cancer Institute*, **102**: 529–37.

Boffetta, P. and Trichopoulos, D. (2008) 'Cancer of the lung, larynx, and pleura', in H. O. Adami, D. Hunter, and D. Trichopoulos, eds, *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. 349–77.

Cordell, H. J. and Clayton, D. G. (2005) Genetic epidemiology 3—genetic association studies. *Lancet*, **366**: 1121–31.

Curado, M. P., Edwards, B., Shin, H. R., Storm, H., Ferlay, J., Heanue,M., and Boyle, P., eds. (2007) *Cancer Incidence in Five Continents, Vol. IX.*. <http://www.iarc.fr/en/publications/pdfs-online/epi/sp160/CI5vol9.pdf>, accessed 6 November 2014.

Ecsedy, J. and Hunter, D. (2008) 'The origin of cancer', in H. O. Adami, D. Hunter, and D. Trichopoulos, eds, *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. 61–85.

Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010) *GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 10*. Lyon: International Agency for Research on Cancer.

Haiman, C. and Hunter, D. (2008) Genetic epidemiology of cancer, in H. O. Adami, D. Hunter, and D. Trichopoulos, eds, *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. 86–108.

Hankinson, S., Tamimi, R., and Hunter, D. (2008) 'Breast cancer', in H. O. Adami, D. Hunter, and D. Trichopoulos, eds, *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. 403–45.

Hung, H. C., Joshipura, K. J., Jiang, R., Hu, F. B., Hunter, D., Smith-Warner, S. A., Colditz, G. A., Rosner, B., Spiegelman, D., and Willett, W. C. (2004) Fruit and vegetable intake and risk of major chronic disease. *Journal of the National Cancer Institute*, **96**: 1577–84.

Hunter, D. J. (2012) Lessons from genome-wide association studies for epidemiology. *Epidemiology*, **23**: 363–7.

Lagiou, P., Adami, H. O., and Trichopoulos, D. (2005) Causality in cancer epidemiology. *European Journal of Epidemiology*, **20**: 565–74.

Lagiou, P., Adami, H. O., and Trichopoulos, D. (2007) Early life diet and the risk for adult breast cancer. *Nutrition and Cancer*, **56**: 158–61.

Lagiou, P., Trichopoulos, D., and Adami, H. O. (2007) 'Epidemiology in the identification of cancer causes', in A. R. Alison, ed.-in-chief, *The Cancer Handbook* (2nd edn). Chichester: John Wiley & Sons, pp. 311–32.

Morrison, A. S. (1992) *Screening in Chronic Diseases* (2nd edn). New York: Oxford University Press.

Potter, J. D. (2005) Vegetables, fruit, and cancer. *Lancet*, **366**: 527–30.

Potter, J. D. and Hunter, D. (2008) 'Colorectal cancer', in H. O. Adami, D. Hunter, and D. Trichopoulos, eds, *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. 275–307.

Schottenfeld, D. and Fraumeni, J. F., eds. (2006) *Cancer Epidemiology and Prevention* (3rd edn). New York: Oxford University Press.

Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians*, **63**: 11–30.

Siemiatycki, J., Richardson, L., Straif, K., Latreille, B., Lakhani, R., Campbell, S., Rousseau, M. C., and Boffetta, P. (2004) Listing occupational carcinogens. *Environmental Health Perspectives*, **112**: 1447–59.

Stuver, S. and Trichopoulos, D. (2008) 'Cancer of the liver and biliary tract', in H. O. Adami, D. Hunter, and D. Trichopoulos, eds, *Textbook of Cancer Epidemiology* (2nd edn). New York: Oxford University Press, pp. 308–32.

Teare, D. M. and Barrett, J. H. (2005) Genetic epidemiology 2—genetic linkage studies. *Lancet*, **366**: 1036–44.

Trichopoulos, D., Adami, H. O., Ekbom, A., Hsieh, C. C., and Lagiou, P. (2008) Early life events and conditions and breast cancer risk: from epidemiology to etiology. *International Journal of Cancer*, **122**: 481–5.

Trichopoulos, D., Li, F. P., and Hunter, D. J. (1996) What causes cancer? *Scientific American*, **275**: 80–7.

Vogelstein, B. and Kinzler, K. W. (2004) Cancer genes and the pathways they control. *Nature Medicine*, **10**: 789–99.

Wacholder, S., Chanock, S., Garcia-Closas, M., and Rothman, N. (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, **96**: 434–42.

Willett, W. C. (2013) *Nutritional Epidemiology* (3rd edn). New York: Oxford University Press.

World Cancer Research Fund and the American Institute for Cancer Research(AICR). (2007) *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*. Washington, DC: AICR.

Chapter 18

Teaching a course in psychiatric epidemiology

Rebecca Fuhrer and Kelly K. Anderson

Introduction to teaching a course in psychiatric epidemiology

The present chapter is intended to serve as a guide for developing a course in psychiatric epidemiology; that is, epidemiological methods applied to the study of psychiatric disorders rather than the epidemiology of specific psychiatric disorders. Methodological issues particularly relevant for these pathologies are emphasized, leaving the delineation of the substantive content of different mental disorders to the instructor's interests and expertise. This course does not need to assume prior training in psychiatry or psychopathology and can be oriented to students at different educational levels. It is expected that students enrolled in this type of course have an interest in medicine, psychology, or public health.

The epidemiology of psychiatric disorders covers the same methods as the epidemiology of other diseases but has additional methodological challenges that need to be made explicit from the outset (Tsuang and Tohen 2002; Susser et al. 2006). These are relevant at the design, analysis, and interpretation phases of epidemiological research. References to epidemiological studies of specific pathologies are included that are particularly illustrative of certain important methodological points. The evolution of research in this application of epidemiology should be reviewed first, followed by methodological issues that have been the focus of research in psychiatric epidemiology. These include measurement (i.e. case definition and case identification), psychometric properties, study design and samples, and theoretical models of environmental and genetic origins of psychopathology. Methodology can be taught in a general course but should be reinforced in the context of this area of application. The subject matter is best communicated in the context of small classes or seminar groups, where there is active teaching for approximately half the session and then interactive exchange based on the taught material and critical analysis of well-selected published work or case studies (Prince et al. 2003).

Psychiatric epidemiology has evolved rapidly in the second half of the twentieth century (Fuhrer and Robins 2007). Earlier work in this field relied on data from mental health providers and facilities, and the limitations of these sources should be explained. Subsequent research endeavours, often community based, frequently defined psychopathology along a continuum from impairment to mental health (Hollingshead and Redlich 1958; Srole et al. 1962). The emphasis then shifted to rates of specific diagnoses in community-based samples using clinical diagnoses at first (Hagnell 1966) and then structured interviewing methods with clinicians (Leighton et al. 1963) and lay interviewers (Eaton and Kessler 1985), with and without computer-assisted technologies. These various methodologies should be contrasted for the student to gain an understanding of the strengths and weaknesses of each approach. Furthermore, studies in children and adolescent samples have increased considerably, with case definition and identification having their own challenges. Finally, longitudinal studies, birth cohort studies, case registries, and administrative databases are increasingly being used for these purposes.

While the number of researchers in psychiatric epidemiology has grown, the 'workforce' remains limited when compared to those working in the fields of cancer or cardiovascular diseases, despite the magnitude of the frequency of the disorders, their impact on quality of life, health services utilization, and economic burden. The level of epidemiological literacy among psychiatrists varies and there are few training programmes that are specifically oriented towards psychiatric epidemiology. The potential usefulness of this approach is often insufficiently addressed during clinical training and, if the students are primarily clinicians, the relevance of epidemiology to clinical practice needs to be discussed.

Teaching objectives

The course can be oriented towards two types of students. The first group would consist of students with a background and training in psychiatry, psychology, or some other mental health specialty. For these students, the instructor would focus more on bringing the methodological issues to bear on the analysis and interpretation of scientific reports. The students should come away from such a course with a sound ability to critically analyse study designs and instruments. Ideally, they should be able to design a study to respond to a specific hypothesis, and be able to discuss the constraints that their study has to incorporate, as well as the resultant limitations of their findings. Although this group would be knowledgeable about psychiatric and psychological disorders, some emphasis should still be placed on diagnostic entities in both

clinical- and community-based samples, the validity of these entities, and measurement reliability (Wing et al. 1981).

The second group of students would consist of general epidemiology students, be they general medical students, established physicians, or other students with public health interests. They would differ from the first group in that they would be lacking expertise in and substantive knowledge of psychiatric disorders; as a result, the instructor would wish to spend some additional time teaching the general notions of psychopathology, different theories of etiology, the symptomatic and diagnostic terminology, and treatment options and efficacy. The students would also need to learn about the prevalence and known risk factors for various mental disorders, in part so that they understood the magnitude and relative importance of this category of health problems. This type of course would need more sessions in order to meet its objectives, with the psychological and psychiatric content taught in a lecture format, complemented with the methodological content as described above.

Teaching content

The general topics of epidemiology, methods, and analysis are covered elsewhere in this volume. Although students may come to such a course with no prior experience or knowledge of psychiatric disorders, the instructor should keep in mind that notions of mental health and mental illness are part of the general culture. Everyone may think they know what 'depression' is; however, this conception of depression will not necessarily be the same as depressive disorders that will be seen by clinicians or identified in epidemiological studies. These general notions need to be discussed and the discussion used as a stepping stone to a scientific approach to the definition of pathology (Wing et al. 1981; Dohrenwend 1990; Wittchen et al. 1999).

A course in psychiatric epidemiology would do well to include some discussion on a general definition of psychopathology, as well as to call attention to topics covered in this chapter, such as case definition, measurement, and study design, that are important to the conception, execution, and interpretation of epidemiological studies in psychiatry (Tsuang and Tohen 2002). If the focus of the course is the epidemiology of psychiatric disorders, presentation of epidemiological data should be reviewed while emphasizing the methodological innovations and/or shortcomings that have contributed to that knowledge (Robins and Regier 1991). An introductory course would do well to include several different types of psychopathology. The following mental disorders, and associated references, provide good teaching material: psychotic disorders such as schizophrenia, for which there is some controversy about frequency and

stability in occurrence across cultures and time (Sartorius et al. 1986; McGrath 2007); affective disorders such as depression, which are frequent (with a rate of occurrence that is hypothesized to be increasing) and have diverse etiologies, including genetic factors (Tsuang and Tohen 2002); and a developmental disorder such as autism, for child psychiatry (Newschaffer et al. 2007).

The issues raised in this chapter are not necessarily unique to psychiatry but partially reflect the current state of knowledge about psychopathology whenever studies are conducted and the concomitant problems of diagnosis and disease classification. Psychiatric epidemiologists often focus on the identification of risk factors, a prerequisite to the determination of causal relationships; many scholars in this field also address secondary prevention via earlier detection and intervention. Once risk factors are identified, other scientific methods of enquiry, as well as epidemiology, can be employed to understand the specific mechanism(s) associated with the increased risk and thereby demonstrate whether a statistical relationship may be shown to be a true cause. However, the instructor should also focus on the fact that knowledge about risk factors can be useful in and of itself, for planning of treatment facilities and for preventive measures.

Measurement issues

For students with knowledge and training in psychiatric disorders, the following topics can be covered by focusing more on the lack of consensus and issues of diagnostic reliability (Susser et al. 2006). These points should be highlighted, as most clinical psychiatrists and mental health providers believe that there is diagnostic agreement and are surprised to observe the contrary when studies demonstrating differences are presented. The instructor must stress that, without diagnostic consensus, all estimates and associations are erroneous at worst, or attenuated at best. Furthermore, it is important to differentiate diagnostic accuracy for clinical and treatment purposes, in contrast to research purposes (Regier et al. 1998).

Case definition

The indices that are usually employed in epidemiology—prevalence and incidence—are based on the assumption that every individual can be classified according to a dichotomous variable: *case* or not. This implies that the boundary between the normal and the pathological can be distinguished with a sufficiently high level of accuracy. In psychiatry the validity of the definitions used is more questionable than in other fields of medicine and reflects the limits of current theories of psychopathology. Unlike other branches of medicine, psychiatry presently has no external criteria that can be used to verify clinical

judgements; a 'gold standard' does not exist. Thus, a definition of a disease or disorder can only be arrived at by consensus.

There are two types of approaches to classification of a case: one is the categorical approach and the other is the dimensional approach. The categorical approach defines psychopathology in a way that results in presumably discrete diagnostic entities. Important examples of the categorical approach are the manuals of disease classification. A widely used system for classifying psychiatric disorders is the *International Classification of Diseases* (ICD; World Health Organization 1992). Psychiatry is the only medical specialty for which a glossary exists that defines each diagnostic category, including guidelines for arriving at a diagnosis. The use of this manual parallels the American experience with the DSM system, which is based on the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association 2013) and is the classification and coding system used primarily in the US. The instructor should stress the multiaxial classification system of the DSM that allows the clinician to evaluate and record information that may be of value in diagnosis, in planning treatment, and in predicting outcome. At present in epidemiology little usage has been made of the axes that evaluate severity and psychosocial functioning.

The use of diagnostic criteria is important because it enhances diagnostic concordance, or reliability, a problem that has always been particularly acute in psychiatry. The advantages and drawbacks of these systems should be addressed and students must be reminded that the usefulness of the categorical approach to case definition in general population studies is uncertain, because the diagnostic nomenclatures were mostly developed from experience with populations that seek care, especially hospital and specialist care. Attention should be drawn to several papers that have evaluated the use of these definitions in community surveys and have found discrepant findings between who is judged to be a 'psychiatric case' (Brugha et al. 1999). Such definitions may not be appropriate for individuals whose configuration of symptoms and signs may differ in terms of duration, co-occurrence, intensity, and impairment. The National Comorbidity Survey in the US attempted to resolve the case definition problem by allowing multiple diagnoses if the criteria were met (Kessler et al. 2003). They obtained high rates of disorders, both for individual disorders and for comorbidity. This could be interpreted as being accurate or that there are multiple manifestations of the underlying constructs and that separating them into several pathologies is artificial.

The challenges associated with changing diagnostic criteria should also be addressed. At the time of writing, the fifth edition of the DSM is currently undergoing field trials, and the eleventh revision of the ICD is in progress (estimated release date 2015). Such revisions to the criteria required for diagnosis

and expansions or restrictions in diagnostic thresholds will substantially impact case ascertainment. For example, each iteration of the DSM has resulted in an increased number of diagnostic categories, more than tripling from the first edition of the DSM (DSM-I) published in 1952 to the fourth edition (DSM-IV) published in 1994 (Double 2002). Given that the revisions are often made in an effort to improve the sensitivity or specificity of a diagnosis, this will have important implications for estimates of the incidence and prevalence of psychiatric conditions. This is especially important to consider for longitudinal studies that span several decades, as well as for the comparability of obtained estimates across different jurisdictions and time periods. Another important point to raise in this context relates to the transferability and validity of Western-culture-based diagnostic systems to a global society. Global mental health studies have seen an important increase in numbers, and a critical perspective of the inferences should be provided to the students.

The dimensional approach defines psychopathology in relative terms, that is, an assessment of the degree of psychopathology or impairment on a continuum. Clinical status is assessed on a particular dimension of psychopathology, such as anxiety, depression, and so forth, or of general psychopathology or psychological well-being. Based on symptom counts, sometimes combined with symptom severity, the addition of item scores calculates a value which varies along a continuum that taps the supposed construct of interest. Similarly, psychiatric disorders may also be conceptualized as illness spectra, whereby similar conditions are grouped together into clusters that may share similar clinical aspects or etiological factors (Andrews et al. 2009). For example, many epidemiological studies will investigate individuals with psychotic disorders, grouping together individuals from diagnostic categories such as schizophrenia, bipolar disorder, delusional disorder, and depression with psychotic features. Although a dimensional or spectral approach may be a closer approximation to clinical and biological realities, its use will have important clinical and policy implications for studies focused on psychiatric interventions or health service provision.

Case identification

Once a definition of a case has been established, the problem remains to identify or measure it in a heterogeneous group of individuals. In order to expose the student to the wide range of instruments available for identifying 'psychiatric cases', the instructor can use the *Handbook of Psychiatric Measures* (Rush et al. 2008), which provides a comprehensive compendium of the many instruments that are available for psychiatric research, Robins' summary of instruments for psychiatric epidemiology (see Tsuang and Tohen 2002), or any of a

number of psychology textbooks, as well as the National Institute of Mental Health (NIMH) website (<http://www.nimh.nih.gov/index.shtml>), which refers to and archives instruments that use a dimensional approach.

The methodology of scale construction should be included in the programme. It should be made clear to the student that clinical observation, theory on psychopathology, and prior research results, combined with consensual processes, lead to the identification and selection of items that are associated with the constructs under study. It is then preferable to use multivariate data analytic techniques to select empirically those signs and symptoms that cluster together best to represent a construct of psychopathology or diagnostic entity. Latent variable modelling methods are useful and pertinent to scale development, and the student should understand what these methods do.

It is of interest to introduce a historical perspective on the evolution and development of psychiatric instruments so that students can better comprehend how and why the instruments that are available now came into being (Eaton and Merikangas 2000; Tsuang and Tohen 2002). The literature during the 1960s and 1970s was replete with references to the lack of reliability in clinical judgement when using the categorical approach for case definition. Studies have shown that diagnostic reliability among psychiatrists, even among those who are experienced and have the same theoretical orientations, was far from satisfactory. The reliability issue was aggravated when cross-national studies were carried out, as shown for adult schizophrenia (Cooper et al. 1972).

Several sources of variance have been identified in the diagnostic process. Patients seen at different examinations may present with different clinical profiles, and symptomatic features are also likely to vary with the evolution of disorders. Clinicians also vary in their methods of obtaining information about a subject, in their skills as interviewers, and as observers of the signs presented by a subject. Moreover, clinicians rely heavily on their personal rules and criteria in the way they use information (i.e. the interpretation of signs and symptoms, and the subsequent organization of these stimuli into a coherent construct or diagnostic category). A useful exercise would include viewing several clinical interviews and then asking students to rate or code the responses and arrive at either a single diagnosis or multiple diagnoses; then the students could compare and calculate diagnostic concordance (Cohen 1960; Fleiss et al. 2003). A similar exercise could use computerized diagnostic interviews.

In their search for solutions to this problematic situation, psychiatric researchers have developed two complementary approaches. One approach is to design a structured method to obtain information about an individual. These structured interviews should result in comparable information for each subject.

It should also be noted that interviews can be observer based or respondent based, that is, either the interviewer judges the sign to be present or the respondent reports the presence or absence of the symptom. Another approach is to establish a set of rules and criteria that help the user arrive at a diagnosis in a consistent fashion. Then, by applying the rules or criteria to the obtained data, one should arrive at the same diagnosis regardless of the interviewer. Again, a historical perspective would be most useful at this point. For example, the Present State Examination (PSE; Wing et al. 1974) and the Schedule for Affective Disorders and Schizophrenia (SADS; Endicott and Spitzer 1978) are two early examples of structured interviews (interviewer based) developed for use by clinicians with psychiatric experience, and now the SCAN (Schedules for Clinical Assessment in Neuropsychiatry; World Health Organization 1994) appears to be its modern successor for epidemiological studies.

The Diagnostic Interview Schedule (DIS; Robins and Regier 1991) was developed for use by lay interviewers in epidemiological surveys and is a respondent-based interview, that is, the data collected are based on what is actually reported by the respondent. Algorithms and rules were developed to apply the diagnostic criteria to the data collected. Other instruments exist and continue to be improved. The Composite International Diagnostic Interview (CIDI), developed under the auspices of the World Health Organization, is the culmination of many years of effort and represents different clinical traditions because it combines elements of the DIS and the PSE and can generate diagnoses based on either the ICD or the DSM traditions. The CIDI has been subjected to intense psychometric scrutiny and exists in many languages. Similarly, the Structured Clinical Interview for DSM (SCID) is a structured diagnostic interview used to assign diagnoses based on the DSM, with the most recent iteration available for DSM-IV (First et al. 1996).

When the dimensional approach is used for defining pathology, case identification habitually entails the use of rating scales (Tsuang and Tohen 2002). Scales can be used by trained observers or may be self-administered instruments. They are usually short symptom checklists that assess a certain level of symptomatology along a given dimension such as depression or anxiety. These types of measures should not be used to arrive at a diagnosis unless they were developed intentionally with that purpose in mind, which is rarely the case. They are usually intended as screening instruments or as methods to assess change in treatment evaluations. Rating scales may be used by psychiatrists or other clinical personnel within clinical settings based on the clinician's assessment of the items. Other rating scales, more appropriate to epidemiological surveys, may be self-report, either completed by the subject or an interviewer, such as the widely used General Health Questionnaire (GHQ; Goldberg 1978), first developed for

primary care patients, or the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff 1977), for use in general population samples.

The students should always be reminded that, when choosing an instrument, it is important to ascertain that it has been evaluated for the context in which its use is planned. Instruments developed for use with patients may be quite irrelevant when used in the general population, and the opposite situation may provide instruments that are completely uninformative when used to discriminate among patients. A set of symptoms observed in a clinical context may not signify the same underlying pathology in community samples. In fact, in both contexts, the constellation may represent valid underlying entities, albeit not the same.

Psychometric issues

The development and use of measures of psychopathology is the subject of numerous books and articles and merits detailed discussion. Key psychometric concepts should be clarified. Students might be referred to the work of Nunnally and Bernstein (1994) as a classic text and guide to other material. There are two basic issues to be addressed in the development of any measure, be it physical or psychological: reliability and validity. Measurement error affects both the estimates of incidence and prevalence and can attenuate or inflate the association between a risk factor and the disorder. Therefore, the link between measurement error and risk assessment should be explained, and the sources of measurement error, their correction, and control should be stressed.

Errors in the data collected may be due to poor reliability and/or inadequate validity of the instruments used. Emphasis should therefore be placed on the necessity to assess these properties in all measures used in a study. Reliability, which refers to the reproducibility or repeatability of measurement, is concerned mainly with chance or random errors due to the observers, the situations, and the instruments. Different ways of assessing reliability should be introduced. Some basic approaches include inter-rater reliability, intra-rater reliability, test–retest reliability, split-half method, and internal consistency. Statistical techniques appropriate for analysing this type of data should be taught: three important examples are Cronbach's alpha, a cumulative item correlation index, for measuring internal consistency (Nunnally and Bernstein 1994); the kappa coefficient as an index of agreement for categorical data (Cohen 1960; Fleiss 2003); and the intra-class correlation coefficient for continuous variables (Fleiss et al. 2003). Multiple measures of the same constructs help reduce measurement error, although they can increase respondent burden.

Reliability is a necessary but not sufficient condition for most forms of validity, and its assessment should be an integral part of all empirical studies. Indeed,

some have argued that current diagnostic systems enhance reliability at the expense of validity. Validity refers to the extent that an instrument in fact measures what it is intended to measure and is not too strongly influenced by systematic error or bias. The different types of validity should be presented and emphasis should be placed on the relative importance of each type in view of a study's objective. Validation efforts should include content validity, construct validity, face validity, and criterion validity (Nunnally and Bernstein 1994).

Content validity refers to the appropriate selection of items from the universe of all observable characteristics of the same latent trait to be included in an instrument. Construct validity refers to the relationship between variables that are measured and the particular trait that they are supposed to tap. Because it links empirical indicators to abstract or theoretical concepts, this type of validity is more difficult to measure. The multimethod, multitrait matrix was one of the first empirical approaches to assess construct validity and may serve as a basis to study this issue (Campbell and Fiske 1959). Factor analysis and other multivariate techniques such as cluster analysis, multidimensional scaling, etc., are central to all types of validity, and their contribution to assessing construct validity should be emphasized. Refinements of statistical techniques proposed four decades ago, that is, the use of latent trait models, have improved instruments in psychiatric epidemiology (Bentler and Stein 1992; Pickles 1998). Criterion validity may be concurrent or predictive, and only the temporal availability of the criterion (present or future) differentiates the two. In psychiatry, criterion validity is usually evaluated in comparison with an 'expert' judgement or a previously validated instrument.

Given the reality that few tests are perfect, one needs to determine the instrument's sensitivity and specificity, and their correlates, the positive and negative predictive value. Students should be reminded of the effect that the prevalence of a disorder has on the predictive values, as well as on the 'true disease' classification. The determination of the aforementioned attributes of any test is discussed in great detail in chapter 24.

Study design

Measures of disease frequency

Some of the problems inherent in the definition and identification of psychiatric cases have already been addressed. The problems of disease onset, duration, relapse, and recurrence are central to the estimation of point prevalence, period prevalence, and incidence rates. Because of the difficulties intrinsic to the assessment of the incidence of psychiatric disorders, for example, identifying the moment of onset and duration), the notion of period and lifetime

prevalence have been employed. Lifetime prevalence is the proportion of individuals in the population who have ever been ill and are alive on a given day. It differs from lifetime risk, which attempts to measure the occurrence of a disease in a birth cohort. When these indicators rely heavily on retrospective assessment of disease occurrence, reliability may be compromised, and estimates of disease rates and risk ratios will be biased. Despite the most recent and sophisticated efforts to design appropriate techniques and instruments, the resulting estimates are probably biased due to recall problems. The importance of information and recall bias, and their effects on risk assessment, should be discussed in detail.

Design

The different types of study design can all be, and have been, used in psychiatric research. The key observational study designs include 'natural experiments', cohort studies, case-control studies, cross-sectional studies, and clinical cohort (or natural history) studies (see Susser et al. 2006 for definitions and examples of their use for psychiatric disorders). Twin studies, adoption studies, and cross-rearing studies can be thought of as special cases of the aforementioned designs. These designs are one way of studying genetic factors but so are family pedigree studies (Risch and Merikangas 1996; Kendler and Prescott 2006). Nowadays, molecular genetics has provided a vast new arena for large-scale case-control studies which seek to detect genomic differences between the case and control groups. Some of these are 'family-based designs'; for example, cases are compared with their biological parents, a method with certain advantages such as in the detection of de novo genetic events (e.g. Sebat et al. 2007).

Numerous studies have used twin and adoption designs, especially in countries that have good case registers and the possibility of linking datasets, such as adoption files in the Scandinavian countries. The choice of a given design depends upon the objective of the study, the rarity of the disease, the putative causes of the disease, and the resources available. However, each design introduces the potential for different types of bias, some of which are particularly thorny in psychiatry, and different problems for the interpretation of the study result. The usual form of the case-control study for risk factor research is used infrequently in psychiatry, particularly because of the issue of recall bias. Denial, repression, negation, screen memories, and rationalization are common defence and coping mechanisms which impede the storage and retrieval of past events in one's psychological life. Reciprocally, individuals who have had personal or familial experience with a mental disorder are often involved in a continuous search for explanatory causes. This factor, combined with persistent feelings of guilt, is likely to falsely influence recall and identification of causes.

Among many examples of the extent of this phenomenon, one may cite Stott's (1958) comparative study, considered a classic, which showed an increased rate of stressful events during pregnancy and thus suggesting a psychological etiology to Down's syndrome.

Samples

Early studies of mental disorders used samples of people treated in psychiatric facilities and, in particular, those admitted to mental hospitals. This raised the issue of two potential sources of bias. First, the types of pathology seen were more likely to be severe cases of psychotic disorders. Second, the characteristics associated with healthcare-seeking behaviour (sex, urbanicity, social class, culture, etc.) were often the variables under study rather than those associated with the pathology. The weight of cultural factors in illness recognition and care should be emphasized within a given country and sociocultural context. These factors vary considerably from one country to another, thereby making cross-cultural comparisons or study replications more difficult. Nevertheless, these facility-based samples may be appropriate, especially for the study of the rarer disorders, such as psychotic disorders or autism (Sartorius et al. 1986).

Choosing community samples is an alternative sampling strategy. The third-generation studies from the 1980s forwards have been carried out on community samples and have used representative samples of the general population in a county, city, region, or country, thereby excluding individuals residing in institutional settings. Alternatively, they have used complete populations of convenience samples, such as the Isle of Wight studies (Rutter 1988), which were conducted on children within a given age range, or the Stirling County study of a total community (Leighton et al. 1963). The Epidemiologic Catchment Area (ECA; Eaton and Kessler 1985) studies carried out in five areas in the US included institutionalized and general population samples, and oversampled certain groups considered at higher risk for psychiatric disorders. Several reviews of psychiatric epidemiology studies and their designs are included in the bibliography and, again, a historical perspective can provide insight for the students (Tsuang and Tohen 2002). These types of community surveys are well suited for collecting information on common mental disorders, such as depression and anxiety. However, they do not often provide useful information about rare disorders, such as schizophrenia. When using lay interviewers for diagnosing complex disorders that are present in less than 1 per cent of the sample, the number of cases detected tends to be small, and the false positives tend to outnumber the true positives. Occasionally, surveys have built in special features to minimize these problems (Susser et al. 2006).

For studies of complex and rare psychiatric disorders, investigators have turned to other approaches, in particular, natural experiments, psychiatric registries, and administrative databases. Natural experiments are usually built around a tragic historic event, although they can also be built around a beneficial event (Costello et al. 2003). The study compares the risk of disorder in people who were exposed to the event with that of those who were not exposed. The most important feature of the design is that the people in the study have little choice about whether or not they will be exposed, thereby minimizing confounding. This design has been used, for example, to link early prenatal starvation to increased risk of schizophrenia in offspring. A first study based on the Dutch Hunger Winter of 1944–45 (Susser et al. 1996), and two subsequent studies based on the Chinese famine of 1960–61, produced concordant results (St Clair et al. 2005; McClellan et al. 2006).

Psychiatric case registries are a source of data for epidemiological studies in those countries that have invested in their development (Perera et al. 2009). In some countries psychiatric registries have been in existence for decades, and the reliability and validity of diagnoses in these registries is better than in surveys. These registries can often identify thousands of cases and can link to other registries to provide information such as birth weight, childhood residence, and school performance. The Danish psychiatric registry has been used, for example, to establish that the risk of schizophrenia is higher among persons born and raised in urban compared with rural areas (Pederson and Mortensen 2006). The Israeli psychiatric registry was used to establish that the risk of schizophrenia is positively correlated with the age of the father at the time of conception (Malaspina et al. 2001). In the highly mobile society that is evolving, local, regional, or even national registries will be compromised with regards to follow-up of the target populations. The student should also learn, however, that good registries are very difficult to establish and that the aforementioned registries are among the few survivors of many attempts to set up registries (Perera et al. 2009). The comprehensiveness of the geographical population and facilities covered should be discussed in view of the cost of registers.

Routinely collected administrative data from health claims is an alternative source of data for epidemiological studies and has recently been growing in popularity, especially in regions that have a publically funded health care system. These databases have an advantage over clinical samples from psychiatric services in that they may additionally capture individuals who are treated in primary care or lost to follow-up after psychiatric referral. They have additional advantages such as the availability of a large number of cases, inclusion of nearly the entire population, and reduced costs.

However, samples obtained from psychiatric case registries and health claim administrative data also pose additional challenges for the definition and interpretation of epidemiologic measures The reliability and validity of the diagnostic and associated data collected on a routine basis need to be examined in order for the student to gain a critical understanding of where and when this data source has and can contribute to the description of the occurrence of certain disorders. Issues such as the scarcity of socio-demographic and clinical information, the lack of diagnostic standardization across professionals (Perera et al. 2009), and limited information on the validity of psychiatric diagnoses (Byrne et al. 2005) should be emphasized. An additional challenge associated with registry and administrative data is the interpretation of any observed trends in studies that attempt to map changes over time. These studies will often be affected by factors such as changing diagnostic systems and the organizational structure of services. The student must also be made aware of the requisite minimum size of the population covered to detect sufficient numbers of the rarer disorders, and the potential for underestimations or biased associations if the inclusion of providers is not comprehensive. For example, a study using an administrative database estimated that approximately 25 per cent of incident cases of psychosis are treated in an outpatient setting only (Jorgensen et al. 2010); therefore, even for rare and complex conditions, the availability of data from outpatient services and primary care could have a substantial impact on observed estimates of incidence and prevalence. Students should also understand that the utility of registers and administrative data is clearly linked to the organization of health care services and the type of reimbursement mechanism and that their existence is dependent on the local or national laws on computer files and data linkage. Discussion could address the type of questions that these data sources can be used to answer and the type of questions for which that data source would be inadequate. Case registers and administrative databases could be used in conjunction with some of the usual study designs and also linked to primary data collected from clinical samples; and this, too, should be elaborated for the students, while reminding them of the limitations.

Assessing students' achievements

The material taught in this type of course is best evaluated with coursework, such as requiring the students to write a mini-proposal, to review a manuscript that would be submitted to a journal that publishes psychiatric epidemiology articles, or to design a study to assess the validity or reliability of an instrument. Short examinations cannot glean the understanding or critical thinking that one would hope the student would have acquired.

Conclusion

This chapter has focused on the important methodological aspects to be covered in a course in psychiatric epidemiology. The orientation has been historical, so that students can learn about what has been done in the past and how to progress and improve in the future. In addition, the methodological challenges that are more acute in this specific branch of epidemiology have been highlighted for the instructor, as the general topics are discussed by other authors in the book. Psychiatric disorders, in fact, cover a wide spectrum of pathologies of diverse origins and different degrees of functional impairment and vary across the lifespan. The instructor should choose the specific pathologies to make his or her methodological points. Clearly, the breadth of ill health spanned by psychiatry and psychology cannot be covered in one epidemiology course, although many of the important research issues can be addressed.

Etiological research of psychiatric disorders has always struggled with the 'nature versus nurture' dilemma. Adoption and twin studies have been used to disentangle the dilemma but results have often been difficult to interpret. Recent advances in genetic epidemiology and molecular biology using linkage analysis, the study of large pedigrees, and genome-wide association studies are promising and tend to support the multifactorial etiologic models of psychiatric disorders. These models include gene–environment interactions, although specific interaction pathways have not fully been established as yet. Students should be exposed to the emerging contributions about the epigenetic influences at critical developmental moments and how they affect gene–environment interaction in the study of normal and pathological development (Meaney and Szyf 2005). Epidemiological research protocols in biological psychiatry should improve the present situation of that discipline.

To date, epidemiological research in psychiatry has helped identify factors associated with certain disorders and has contributed to improving diagnostic reliability and validity. Future research endeavours need to incorporate multidisciplinary approaches, including biological markers, social and genetic risk, and protective factors. The contributions of developmental psychology and sociology to the understanding of social support, as well as chronic and acute stress, as factors in the occurrence of illness, psychological and physical, need to be understood and encouraged. Research on the etiology of specific disorders will only progress once hypotheses and models of the causal and mediating factors, their relationships to the disorder and to each other, are made explicit, tested, and rejected (or not). Elaborating such causal relationships, designing the appropriate studies, and employing the proper analytical procedures, is now the goal of the next generation of research in psychiatric epidemiology.

References

American Psychiatric Association. (2013) *Diagnostic and Statistical Manual of Mental Disorders* (5th edn). Washington, DC: American Psychiatric Association.

Andrews, G., Goldberg, D. P., Krueger, R. F., Carpenter, W. T., Hyman, S. E., Sachdev, P., and Pine, D. S. (2009) Exploring the feasibility of a meta-structure for DSM-V and ICD-11: Could it improve utility and validity? *Psychological Medicine*, **39**: 1993–2000.

Bentler, P. M. and Stein, J. A. (1992) Structural equation models in medical research. *Statistical Methods in Medical Research*, **1**: 158–81.

Brugha, T. S., Bebbington, P. E., Jenkins, R., Meltzer, H., Taub, N. A., Janas, M., and Vernon, J. (1999) Cross validation of a general population survey diagnostic interview: a comparison of the CIS-R with SCAN ICD-10 diagnostic categories. *Psychological Medicine*, **29**: 1029–42.

Byrne, N., Regan, C., and Howard, L. (2005) Administrative registers in psychiatric research: a systematic review of validity studies. *Acta Psychiatrica Scandinavica*, **112**: 409–14.

Campbell, D. T. and Fiske, D. W. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**: 81–105.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**: 37–46.

Cooper, I. E., Kendell, R. E., Gurland, B. J., Sharpe, L., Copeland, J. R. M., and Simon, R. J. (1972) *Psychiatric Diagnosis in New York and London*. Oxford: Oxford University Press.

Costello, E. J., Compton, S. N., Keeler, G., and Angold, A. (2003) Relationships between poverty and psychopathology: a natural experiment. *Journal of the American Medical Association*, **290**: 2023–9.

Dohrenwend, B. P. (1990) The problem of validity in field studies of psychological disorders revisited. *Psychological Medicine*, **20**: 195–208.

Double, D. (2002) The limits of psychiatry. *British Medical Journal*, **324**: 900–4.

Eaton, W. W. and Kessler, L. G., eds. (1985) *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program*. London: Academic Press.

Eaton, W. W. and Merikangas, K. R. (2000) Psychiatric epidemiology: progress and prospects in the year 2000. *Epidemiologic Reviews*, **22**: 29–34.

Endicott, J. and Spitzer, R. L. (1978) A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, **35**: 837–44.

First M.B., Spitzer R.L., and Gibbon M.B. (1996) *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version*. Washington, DC: American Psychiatric Press.

Fleiss, J. L., Levin, B., and Paik, M. C. (2003) *Statistical Methods for Rates and Proportions* (3rd edn). New York: John Wiley & Sons.

Fuhrer, R. and Robins, L. (2007) 'The epidemiological study of mental disorders since the beginning of the twentieth century', in W. W. Holland, J. Olsen, and Charles du V. Florey, eds, *The Development of Modern Epidemiology: Personal Reports from Those Who Were There*. Oxford: Oxford University Press.

Goldberg, D. P. (1978) *Manual of the General Health Questionnaire (GHQ)*. Windsor: NFER-Nelson.

Hagnell, O. (1966) *A Prospective Study of the Incidence of Mental Disorders: the Lundby Report*. Stockholm: Svenske Bokforlaget, Norstedts.

Hollingshead, A. B. and Redlich, F. C. (1958) *Social Class and Mental Illness*. New York: Wiley.

Jörgensen, L., Ahlbom, A., Allebeck, P., and Dalman, C. (2010) The Stockholm non-affective psychoses study (SNAPS): the importance of including outpatient data in incidence studies. *Acta Psychiatrica Scandinavica*, **121**: 389–92.

Kendler, K. S. and Prescott, C. A. (2006) *Genes, Environment, and Psychopathology*. New York: Guilford Press.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, J., Walters, E. E., and Wang, P. S. (2003) The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association*, **289**: 3095–105.

Leighton, D. C., Harding, J. S., Macklin, D. B., Hughes, C. C., and Leighton, A. H. (1963) Psychiatric findings of the Stirling County study. *American Journal of Psychiatry*, **119**: 1021–6.

Malaspina, D., Harlap, S., Fennig, S., Heiman, D., Nahon, D., Feldman, D., and Susser, E. S. (2001) Advancing paternal age and the risk of schizophrenia. *Archives of General Psychiatry*, **58**: 361–7.

McClellan, J. M., Susser, E., and King, M. C. (2006) Maternal famine, de novo mutations, and schizophrenia. *Journal of the American Medical Association*, **296**: 582–4.

McGrath, J. (2007) The surprisingly rich contours of schizophrenia epidemiology. *Archives of General Psychiatry*, **64**: 14–15.

Meaney, M. J. and Szyf, M. (2005) Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience*, 7: **103**–23.

Newschaffer, C. J. et al. (2007) The epidemiology of autism spectrum disorders. *Annual Reviews of Public Health*, **28**: 235–58.

Nunnally, J. C. and Bernstein, I. H. (1994) *Psychometric Theory* (3rd edn). New York: McGraw Hill Higher Education.

Pedersen, C. B. and Mortensen, P. B. (2006) Are the cause(s) responsible for urban—rural differences in schizophrenia risk rooted in families or in individuals? *American Journal of Epidemiology*, **163**: 971–8.

Perera, G., Soremekun, M., Breen, G., and Stewart, R. (2009) The psychiatric case register: noble past, challenging present, but exciting future. *British Journal of Psychiatry*, **195**: 191–3.

Pickles, A. (1998) Psychiatric epidemiology. *Statistical Methods in Medical Research*, **7**: 235–51.

Prince, M., Stewart, R., Ford, T., and Hotopf, M. eds, (2003) *Practical Psychiatric Epidemiology*. Oxford: Oxford University Press.

Radloff, L. S. (1977) The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, **1**: 385–401.

Regier, D. A., Kaelber, C. T., Rae, D. S., Farmer, M. E., Knauper, B., Kessler, R. C., and Norquist, G. S. (1998) Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Archives of General Psychiatry*, **55**: 109–15.

Risch, N. and Merikangas, K. R. (1996) The future of genetic studies of complex human diseases. *Science*, **273**: 1516–17.

Robins, L. N. and Regier, D. A. (1991) *Psychiatric Disorders in America*. New York: The Free Press.

Rush, A. J., First, M. B., and Blacker D., eds. (2008) *Handbook of Psychiatric Measures* (2nd edn). Washington, DC: American Psychiatric Publishing.

Rutter, M., ed. (1988) *Studies of Psychosocial Risk: the Power of Longitudinal Data*. Cambridge: Cambridge University Press.

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., and Day, R. (1986) Early manifestations and first-contact incidence of schizophrenia in different cultures: a preliminary report on the initial evaluation of the WHO collaborative study on determinants of outcome of severe mental disorders. *Psychological Medicine*, **16**: 909–28.

Sebat, J. et al. (2007) Strong association of de novo copy number mutations with autism. *Science*, **316**: 445–9.

Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., and Rennie, T. A. (1962) *Mental Health in the Metropolis: the Midtown Manhattan Study*, vol. I. New York: McGraw Hill.

St Clair, D. et al. (2005) Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959–1961. *Journal of the American Medical Association*, **294**: 557–62.

Stott, D. H. (1958) Some psychosomatic aspects of causality in reproduction. *Journal of Psychosomatic Research*, **3**: 42–55.

Susser, E., Neugebauer, R., Hoek, H. W., Brown, A. S., Lin, S., Labovitz, D., and Gorman, J. M. (1996) Schizophrenia after prenatal famine: further evidence. *Archives of General Psychiatry*, **53**: 25–31.

Susser, E., Schwarz, S., Morabia, A., Bromet, E. J., Begg, M. D., Gorman, J. M., and King, M. C. C. (2006) *Psychiatric Epidemiology: Searching for the Causes of Mental Disorders*. New York: Oxford University Press.

Tsuang, M. T. and Tohen, M., eds. (2002) *Textbook in Psychiatric Epidemiology* (2nd edn). New York: Wiley-Liss.

Wing, J. K., Bebbington, P., and Robins, L. N. (1981) *What is a Case? The Problem of Definition in Psychiatric Community Surveys*. London: Grant McIntyre.

Wing, J. K., Cooper, J. E., and Sartorius, N. (1974) *The Measurement and Classification of Psychiatric Symptoms*. Cambridge: Cambridge University Press.

Wittchen, H-U., Ustun, T. B., and Kessler, R. C. (1999) Diagnosing mental disorders in the community: a difference that matters. *Psychological Medicine*, **29**: 1021–7. World Health Organization. (1992) *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.

World Health Organization. (1994) *SCAN: Schedules for Clinical Assessment in Neuropsychiatry, Version 2.0*. Geneva: Psychiatric Publishers International.

Chapter 19

Neurologic diseases

C. A. Molgaard, A. L. Golbeck,
and John F. Rothrock

Introduction to neurologic diseases

Neuroepidemiology has traditionally been defined as the study of the distribution and determinants of neurologic diseases and injuries. A number of textbooks in the area have been published in the last few decades (Anderson and Schoenberg 1991; Molgaard 1993; Gorelick and Alter 1994; Batchelor and Cudkowicz 1999; Nelson et al. 2004) The journal *Neuroepidemiology*, originally edited by Bruce Schoenberg, was established in 1982 and is currently thriving under the leadership of Valery Feigin of Auckland University in New Zealand.

Neuroepidemiology, like all forms of epidemiology, depends on the correct classification and ascertainment of outcomes, as well as understanding the complexities of exposure along the life course. Tanner and Ross (2004), in their chapter 'Neuroepidemiology: fundamental considerations', focus on disease-specific challenges in the conduct of neuroepidemiologic research, as well as the need to formulate a well-articulated research question, as this question will guide the formation of the specific aims of the study, the specific hypotheses to be addressed, the selection of the population for study, the selection of outcome measures, and the analysis plan. Many of the disease-specific challenges to neuroepidemiology, such as varying diagnostic criteria over time, the need for post-mortem examination for definitive diagnosis, the relative rarity of the diseases, the need for a precise time of disease onset, and a long latency period, can be partially addressed by the use of disease registries and records-linkage systems.

Here the epidemiological research of the many epidemiologists and statisticians involved in working at the Mayo Clinic's Rochester Epidemiology Project should be noted (Kurland and Molgaard 1981). The availability of a complete and complex clinically based records-linkage system (patient records indexed by diagnosis back to 1935) for Olmsted County, Minnesota, has allowed

population-based uses of multiple sources of information on neurologic outcomes with a high degree of diagnostic accuracy. This model was recently exemplified in the European context by the research of Hallas et al. (2008), who used data gathered by the Danish governmental organization Statistics Denmark to focus on stroke and myocardial infarction. In this study, the research team used a national prescription database with comprehensive recordings of all prescriptions filled in Denmark since 1995 and linked to a number of other health databases to assess the toxicity of a dietary drug. Health databases such as the Danish National Registry of Patients (all individual hospital discharges since 1977), the Prescription Database of the Danish Medicines Agency (data on all prescriptions since 1995), the Danish Register of Deaths, and the Danish Person Registry were linked for analysis using a unique identifier in a study that focused on a cohort of 298,848 persons.

The teaching of this subfield of epidemiology, also a subfield of neurology, is important because of the large public health burden which neurologic diseases and injuries represent at the population level. The characteristic social epidemiology of common neurologic disorders has been addressed by both Molgaard (2002) and Rothrock and colleagues (Rothrock, Lyden, Brody, et al. 1993; Zweifler et al. 2002). The latter specifically noted that, along with prevalence and incidence, the etiologies and clinical manifestations of the most common neurologic disorders may vary according to the general health and health-related behaviour of the individual populations examined.

For example, in south Alabama, where smoking, obesity, diabetes, and untreated hypertension are relatively common, the annual incidence of stroke is proportionately higher than it is in areas of the US where those risk factors are less often present. The causes of stroke in south Alabama correspondingly tend towards the 'traditional', with atherosclerosis typically promoting the ischaemic variety and hypertension promoting intracerebral haemorrhage. In contrast, investigators evaluating the etiologies of stroke in San Diego found relatively lower incidences of atherosclerotic stroke and hypertensive intracerebral haemorrhage and relatively higher incidences of 'ischaemic stroke of unknown cause' and intracerebral haemorrhage induced by 'recreational' use of sympathomimetic drugs. It follows that the success or failure of interventions intended to reduce stroke incidence, morbidity, and mortality may depend in part on how well those interventions match up with the underlying causes of stroke in the target population.

While the etiology, prevalence, and incidence of migraine, another common neurologic disorder, may remain constant from one population to another, the clinical characteristics and relevant behaviour of migraineurs within those populations may differ considerably. Chronic migraine is a common and costly

clinical variant of migraine that afflicts approximately 2 per cent of the general population, a prevalence that differs little according to the particular population surveyed (Bigal et al. 2008). Patients with chronic migraine commonly overuse symptomatic medication, and such overuse may exacerbate the underlying primary headache disorder and cause the patients to experience so-called medication overuse headache (MOH) (Silberstein et al. 2005). The incidence of MOH varies significantly according to geographic region, and the specific medications overused also may vary. As to the latter, overuse of non-prescription 'simple' analgesics or relatively migraine-specific therapies such as the triptans may prevail in one region, while overuse of opiates/opioids or butalbital-containing compounds may dominate in another. As the susceptibility of MOH to treatment is in part a function of the specific medication(s) being overused, a failure to identify this epidemiologic pattern may inhibit the successful management of chronic migraine within the population under consideration.

The importance of neuroepidemiology also lies in that it spans a large range of medical, social, botanical, geographic, and behavioural sciences. The result has been the sharing of interests and concepts across disciplinary boundaries in the common language of epidemiology and in the elucidation of new notions of disease risk and disease progression.

The classic example of the latter in neuroepidemiology was the work of Carleton Gadjusek, beginning in the late 1950s, on a disease called kuru found among the Fore tribe of the New Guinea Highlands. For his research on the mechanisms for the origin and dissemination of a neurologic disease caused by a slow virus, Gadjusek received the Nobel Prize in Medicine in 1976, and he is the only epidemiologist to have done so. His research was unique in that it integrated anthropology, neurology, epidemiology, genetics, and statistics in solving a highly different type of neurologic disease outbreak (Molgaard 1981). Among many others, neuroepidemiologists such as Leonard Kurland of the Mayo Clinic and statisticians such as R. A. Fisher of Rothhamsted Experimental Station contributed to the elucidation of this conceptual breakthrough.

Research and teaching models of neuroepidemiology have changed through time. Classically, the paradigm has been a population laboratory approach to neurologic diseases and injuries that emphasized the etiologic importance of geographic isolates. A geographic isolate is defined as a population that may be inbred, remote, isolated, or otherwise distinct, and is of interest because a high incidence of a new disease or a high prevalence of a common disease in such a population may reveal etiologic clues that would otherwise be obscure (Kurland 1978; Kurtzke and Hyllested 1979). This paradigm originally came out of the

Table 19.1 Classical model of neuroepidemiology used in teaching and research.

Disease surveillance emphasized	Geographic isolates used to elucidate etiology
Methodologic in orientation	Individual and social risk factors used to describe disease patterns
'Slow' virus and 'slow' toxin models of exposure accepted	Collaboration across disciplines

field of geographic pathology. Other specific research activities then evolved within this context (Table 19.1). These include

- population surveillance—this involves tracing trends in neurologic diseases and injuries (e.g. the development of stroke registries for defined populations; Rothrock, Lyden, and Brody 1993);
- the development of improved research methodologies (e.g. refining and validating research instruments to screen populations for Alzheimer's disease (Hough et al. 1993; Prince 1998), or the use of population-based records-linkage systems, such as that of the Mayo Clinic, to study neurologic disease (Kurland and Molgaard 1981));
- social and behavioural applications—research is oriented to individual and group risk factors for neurologic injury and disease (e.g. studies of the social and behavioural correlates of epilepsy, whether as risk factors per se or barriers to diagnosis and treatment; Paschal et al. 2005, 2007); and
- international applications (e.g. research on the global patterns of neurologic disease and injury of such illnesses as multiple sclerosis, with its highly characteristic north—south gradient of higher to lower incidence (Kurtzke et al. 1993; Kurtzke 2000), or of the controversies surrounding beef consumption and Creutzfeldt–Jacob disease (CJD) in Great Britain and other countries of western Europe (Molgaard and Golbeck 1992; Pollack 1999)).

However, this paradigm has evolved (Goldman and Kodura 2000; Tilson 2000). There have been major theoretical advances and the addition of new exploratory concepts from environmental health and toxicology that deserve attention in a class on neuroepidemiology, or in a segment on neurologic diseases that is a part of a general class on epidemiology (Box 19.1). Most of these concepts relate to the programming hypothesis, which holds that critical windows of exposure exist for children in terms of developmental toxicants and that these early exposures can lead to late neurodegenerative effects.

Many etiologic exposures (including infections) are now understood to be in utero, involving both acute and delayed effects and exhibiting both structural and functional defects (Adams et al. 2000). In essence, the relationship between

Box 19.1 Slow-exposure model of neuroepidemiology used in teaching and research

- Disease surveillance still a priority but based increasingly on web-based activities
- Increased methodologic orientation to genetics, nutrition, and developmental neurotoxicity
- Geographic isolates of less importance
- Individual and social risk factors more clearly related to major theories of psychology, sociology, and anthropology, as well as behavioural toxicology
- Increased collaboration across disciplines
- 'Slow' models of exposure extended to in utero and postnatal periods
- 'Chain of causation' requiring evaluation of new theories, including current animal models relating to critical periods of environmental exposures in utero and during development

environmental exposures and structural and neurobehavioural effects has become more complicated and more common, and develops much earlier than previously suspected (Gluckman and Hanson 2005).

Both approaches involve a continuing evolution of our training paradigm, from (1) classical neuroepidemiology to (2) a more exposure-focused neuroepidemiology including 'slow' models of exposure in the in utero and postnatal periods, to a clinical neuroepidemiology with a greater focus on diagnosis, treatment, and prognosis. The move historically in this subfield of epidemiology is then one from primary prevention in classical neuroepidemiology to secondary and tertiary prevention in a more clinically based neuroepidemiology, a move that would appear to be shared by European and American epidemiologists. However, the American model emphasizes the attainment of a master's degree in public health (MPH), health services, or a related field. Neuroepidemiology is a common elective course in MPH curricula.

Teaching objectives

The overall teaching goal of this orientation to neuroepidemiology is to create an understanding of the complexity of the issue of exposure. Environmental exposures not only may be very slow but may occur very early in the lifespan

and be deeply modified in impact by the genome. A good analogy comes from cardiovascular disease, where intrauterine growth restriction and low birth weight are known to be associated with a higher risk for hypertension, coronary artery disease, and diabetes in adult life (Osmond and Barker 2000). Practically, this means that the student needs to gain an appreciation for the fact that each particular type of defect is associated with a specific window of vulnerability and that developmental toxicants can include true teratogens, embryotoxins, and foetotoxins as they apply to neurologic structure and process (Adams et al. 2000).

Learning objectives

For practical classroom purposes, it is important to emphasize key concepts of programming with specific human diseases as well as animal models. While it is clear that extrapolating from animal data to humans is often fraught with peril, it also often provides a means of thinking in new ways about exposure and outcome that can be highly stimulating. To begin, the teacher needs to achieve buy-in from students on the following learning objective:

Learning objective 1—A good neuroepidemiologic research strategy is team-based and is devised from previous empirical research that is a combination of (a) clinical studies, (b) animal studies, and (c) epidemiologic studies.

Accomplishing this buy-in can be difficult. However, there are multiple examples in the medical and public health literature that can drive the point home. One of the most useful approaches is to illustrate the learning objective with major public health victories that are clear-cut. The eradication of smallpox, for example, is always good in general, while a more neuroepidemiologic example relates to polio and the teamwork involved in development of the Salk and Sabin vaccines (Oshinsky 2005).

Learning objective 2—Theoretical epidemiology needs to include models that are beyond mere statistical association, instead including a better understanding of molecular and cellular processes that underlie induction of disease and variable latency periods.

In neuroepidemiology, we may find a number of examples of this in terms of vulnerable periods in the development of the central nervous system (brain and spinal cord). The vulnerability of the central nervous system is dependent on two issues (Rice and Barone 2000). First, does an agent (or its active metabolites) reach the developing nervous system; second, what was the period of exposure? Here we are reaching an overlap with much of the work in molecular epidemiology from the early 1990s (Schulte and Perera 1993), and recent advances summarized by Foxman (2012). The difference is that our knowledge

of the developing nervous system has improved to the point that there is recognition that, first, the developing nervous system is qualitatively different from its adult analogue and, second, that exposure before or after an organ develops is less perturbing in a system's sense than if exposure occurs during development of the organ per se (Rice and Barone 2000).

Learning objective 3—Developmental neurotoxicity may have small effects on the individual but, if the impact occurs across an entire population as well as across the lifespan of individuals, the societal impact in terms of public health burden can be massive.

The cause of most neurodevelopmental disabilities—perhaps more than 75 per cent—is unknown. Those that we are beginning to understand in terms of etiology to some extent include dyslexia and mental retardation. In terms of magnitude, approximately 3 to 8 per cent of all births in the US (4 million babies a year) are affected by these disorders. Yet another 1 million children in the US still suffer from the effects of elevated blood lead levels with its attendant impact on neurobehavioural functioning and intellectual activity (Weiss and Landrigan 2000).

Students need to come away from this course understanding that even a decrement in function that is within the normal range, if exposure is widespread in a society, can have an enormous impact. An example related to lead is loss of IQ. Even a small loss of perhaps five points in many lead-exposed individuals would shift the tails of the normal distribution dramatically—doubling the number of those with IQs below 70 and who thus require greater educational and social resources from a society (Rice and Barone 2000). There would also be a reduction of the numbers of those with high IQs (greater than 130). Considerable research has been carried out in this area by Hernstein and Murray (1996). A historical example of this process that students often enjoy is the theory that one of the reasons for the decline and fall of the Roman Empire was the ubiquitous nature of lead in Roman drinking vessels and plates during the late empire, with social decay and lack of leadership being associated with widespread lead poisoning and intellectual impairment (Gilfillan 1965; Lewis 1985; Hernberg 2000). The relationship between low IQ and current global iodine deficiency should also be noted (Anderson et al. 2005).

Teaching content, method, and format

Content should be of two types. The first is to orient the student to a number of key concepts related to neurogenesis and programming, including neural proliferation, migration of recently proliferated cells, differentiation of neuroblasts to a terminal phenotype, synaptogenesis, gliogenesis, myelination, apoptosis, neurotransmitters, and neurotrophic signalling (Rice and Barone 2000).

The second type of content should then relate to eight key questions concerning risk estimation in neuroepidemiology. These are as follows (Adams et al. 2000):

1. In terms of vulnerability in the development of the nervous system, which time periods carry the greatest risk?
2. Are there cascades of developmental disorders in the nervous system?
3. Can critical windows of vulnerability suggest the most susceptible subgroups of children by geographic area, socio-economic status, race, sex, or other demographic/social parameters?
4. What data gaps exist regarding the endpoints of an environmentally altered nervous system?
5. What are the best ways to examine exposure—response relationships and estimate exposures during critical periods of development?
6. What other exposures during development may interact with and alter exposures of concern?
7. How well do laboratory animal response data parallel human response data?
8. How can this type of data be used in risk assessment and public health?

Once these concepts have been introduced, each student should then be assigned a specific area of neurogenesis (e.g. myelination). The student then needs to research this area in terms of the scientific literature in neurology, epidemiology, and neurotoxicology for specific disease examples that are related to this aspect of development.

The research effort should end with the student presenting to the entire class what was found. The presentation should try to follow the eight conceptual areas (vulnerable time periods, cascades of disorders, critical windows of exposure, endpoints, estimating exposures, interacting exposures, parallels from animal laboratory studies, and risk assessment in public health) outlined by Adams et al. (2000) as closely as possible but be specific to one disease. Following the presentation and discussion in a seminar-type style, the student should write a report summarizing his/her findings for the disease noted during the research effort and include feedback from the rest of the class and the instructor regarding the one specific disease presented for discussion to the class. This discussion will lead to the final project of the class, where the student designs a research project of the disease in question.

If we return to the myelination example, a student might be encouraged to examine what we currently know about critical periods of myelin development and possible neurotoxic environmental hazards in terms of the unique north—south distribution of multiple sclerosis. We now know that myelination begins

in the late prenatal period (around months 5 or 6) and continues during the postnatal period for different systems up to the age of 10 years. We also know that, when people migrate, age at migration determines multiple sclerosis risk. Those who leave a high-risk northern latitude after age 15 take the higher risk for developing this disease with them to the low-risk southern latitudes, and vice versa. Is the completion of myelination at age 10 related to this unique risk factor profile? What environmental risk factors are suspect, given the period of vulnerability for myelination and its pathologic involvement in multiple sclerosis? What types of research design might be most productive for studying such risk factors? What possible sources of bias and confounding exist in such a study?

During the seminar section of the class, the instructor should revisit the three key learning objectives repeatedly. Other topics such as childhood development in contaminated urban and rural settings (Guillette 2000) or the impact of environmental agents (e.g. ethanol, barbiturates, etc.) in triggering massive apoptotic neurodegeneration during the last trimester of pregnancy and first several years after birth (Olney et al. 2000) can be introduced at this time.

Digital learning

The journal *Distance Education Learning* is an invaluable tool in creating classes using digital learning. The most recent issue focused on issues of online learning test-takers, the ways digital online learning can lead to deeper learning, and other aspects of the virtual classroom (Magna Publications 2013). The challenges of the online classroom are far surpassed by the advantages of reaching a diverse audience at a distance in an asynchronous fashion.

Digital learning is being used with success in both neurology and epidemiology. Digital learning in neurology has already taken on a variety of shapes and forms, including—but not limited to—continuing medical education that involves sharing neurology knowledge to improve patient care; neurology clerkships that focus on weekly case lectures and answers; courses that offer neurology instruction complementing clinical training; short courses that help general practitioners to update their knowledge and skills about patients with neurological disorders; courses that add to the neurology knowledge of practitioners of alternative medicine; and mini courses for K–12 students focusing on neuroscience, careers in neuroscience, and how to stay brain healthy. Similarly, digital learning in epidemiology has been developed for a variety of goals/objectives, target audiences, levels, prerequisites, intensities, types of credit, and classes across the national and international scene.

A first digital learning course in neuroepidemiology (Public Health 595: Neuroepidemiology—Instructors Curtis Noonan and Craig Molgaard) was

taught within an online master's degree programme in public health at the University of Montana in spring 2007, with an enthusiastic response from learners. This course was implemented using the Blackboard e-Education platform, with course competencies similar to those of a face-to-face MPH neuroepidemiology course previously taught by Molgaard in medical school and graduate school of public health venues. Following best practices in digital education, the course was developed in its entirety before being brought online.

Learning was structured into self-study and readings, discussion board, and a project, and it incorporated considerable interaction, a primary ingredient for quality of any digital learning course. In addition, the final section of the course in 2007 was called 'Community Response to Environmental Disasters with Neurologic Endpoints'. This was a day-long symposium co-sponsored by the Mansfield Library and the School of Public and Community Health Sciences at the University of Montana, as well as the Japan Foundation and others, focusing on the mercury-poisoning disaster of Minamata Bay, Japan (and parallels to the Libby, Montana, asbestos-vermiculite Superfund site). Guest speakers from the Kumamoto Prefect Office and the Libby site were brought in, and students from the neuroepidemiology class were invited to attend if possible. The symposium was filmed for those who could not attend, and CDs were mailed to all in the web-based neuroepidemiology class.

The most recent Neuroepidemiology class was given using the Moodle online platform (see the current syllabus in Box 19.2). This class was given during the spring semester 2013 by Molgaard and Kathryn Fox of the School of Public and Community Health Sciences at the University of Montana. It emphasized the toolkit of molecular epidemiology for use by neuroepidemiologists.

Assessing students' achievements

Styles of assessment depend to some degree on whether a neuroepidemiology class is taught face to face or by digital techniques. Nevertheless, there are some commonalities. Contributions to discussion can be assessed in terms of quality, whether in person in the classroom or by discussion thread over the web. Presentations can be verbal or in writing, with appropriate feedback from other students and instructors. Exams can be carried out through digital procedures as well as in classroom settings, especially if timed approaches are used. Final class projects can also focus on the integration of concepts and methods of neuroepidemiology in either format. For our most recent digital neuroepidemiology class, the final project focused on defining a research question and hypothesis and preparing a research proposal using appropriate literature, methods, and data management and analysis procedures. This included

Box 19.2 PUBH 512 Neuroepidemiology

Instructors: Craig A. Molgaard, Ph.D., M.P.H., Professor and Chair
Kathryn Fox, J.D.
School of Public and Community Health Sciences
College of Health Professions and Biomedical Sciences
The University of Montana
Format: Online seminar with Moodle. Course procedures and expectations are those normally used in the online MPH program at the University of Montana.

Required Texts:

1. Nelson, Lorene, et al. Neuroepidemiology. New York: Oxford University Press, 2004.
2. Sacks, Oliver. The Island of the Colorblind. New York: Vintage Books, 1996.
3. Sacks, Oliver. Hallucinations. New York: Vintage Books, 2012.
4. Sorbo, Kevin. True Strength. Cambridge, Mass.: Da Capo Press, 2012.
5. Ochinsky, Peter. Polio: An American Story. Oxford: Oxford University Press, 2005.
6. Foxman, Betsy. Molecular Tools and Infectious Disease Epidemiology. New York: Elsevier-Academic Press, 2012.

Other readings as assigned will be available online via the Moodle classroom.

Course Description:

This course focuses on the epidemiology of neurological diseases. Students will learn about special considerations researchers incorporate into studying neurological diseases. In addition, students will write a literature review on a particular disease and create their own study about that disease with these considerations in mind.

Online Format: Readings from the textbooks and discussion questions from the texts and the instructor will form the basis of weekly class postings on the online discussion board. The course will conclude with a final term paper.

Course Evaluation: Weekly postings to the discussion board will make-up class participation, which constitutes 30 per cent of the course grade. The

Box 19.2 PUBH 512 Neuroepidemiology *(continued)*

mid-semester literature review will constitute 30 per cent of the overall grade and the final term paper will constitute 40 per cent of the overall grade.

Course Schedule:

Week 1
The Island of the Colorblind by Oliver Sachs, Book I
Week 2
The Island of the Colorblind by Oliver Sachs, Book II
Week 3
Fundamental Considerations of Neuroepidemiology
Study Design, Measures of Effect, and Sources of Bias
Readings:

- Nelson, chs. 1–2
- Tsuang, Debby, et al. 'GBA mutations increase risk for Lewy body disease with and without Alzheimer disease pathology.' Neurology. 79.19(2012): 1944–1950.
- Molgaard, C. and Golbeck, A. 'Mad Cows and Englishman: bovine spongiform encephalopathy (BSE).' Neuroepidemiology. 9(1992):285–286.
- Washington Post Article on Nobel Prize winner D. Carleton Gadjusek: J. Gilllis and J. Spinner. 'A Life of Rare Purpose and Passion.' Washington Post, Friday April 26, 1996.

Assignment:

- Choose a disease chapter from the Nelson text, chs. 5–15, on which you will focus your midterm and final projects and let Kathryn know which disease you have chosen (no two classmates may choose the same disease).
- Start working on literature review, due Week 6.

Week 4
Measurement and Analysis
Genetic Epidemiology of Neurologic Disease
Readings:

- Nelson, chs. 3–4
- Baranzini, Sergio E. and Dorothee Nickles. 'Genetics of multiple sclerosis: swimming in an ocean of data.' Current Opinion in Neurology. 25.3(2012):240–245.

Box 19.2 PUBH 512 Neuroepidemiology *(continued)*

- Chen, X., et al. 'ApoE and CYP2D6 polymorphism with and without parkinsonism-dementia complex in the people of Chamorro, Guam.' Neurology. 47.3(1996):779–84.
- Kurtzke, J.F. and Heltberg, A. 'Multiple sclerosis in the Faroe Islands: an epitome.' J. Clin. Epidemiol. 54(2001):1–22.

Week 5

Hallucinations by Oliver Sacks and Test of Strength by Kevin Sarbo, winner of the American Brain Foundation's 2013 Public Leadership in Neurology award.

Week 6

Current Neurology Issues

Readings:

- Browse a recent (past 5 months) issue or two of Neurology Today, Neurology, Neuroepidemiology, Lancet Neurology, etc.

Assignment:

- Post a discussion board topic discussing one article that you found particularly interesting, why, and identify strengths and weaknesses with the arguments and research strategies employed in the article
- While each student must post a different article with a few comments about the article, students need only respond to two other postings.

Week 7

Disease discussions begin

Readings:

- Relevant chapter chosen by classmate
- Orlinsky, Polio: An American Story

Assignment:

All literature reviews due

Critique classmate's literature review on his or her chosen disease

Week 8

Readings:

- Relevant chapter chosen by classmate
- Foxman, Molecular Tools, Chapters 1–3 (Introduction and Historical Perspective, How Molecular Tools Enhance Epidemiological Studies, Applications of Molecular Tools to Infectious Disease Epidemiology)

Assignment:

Critique classmate's literature review on his or her chosen disease

Box 19.2 PUBH 512 Neuroepidemiology *(continued)*

Week 9

Readings:

- Relevant chapter chosen by classmate
- Foxman Molecular Tools, Chapters 4–6 (A Primer of Epidemiologic Study Designs, A Primer of Molecular Epidemiology, Molecular Tools)

Assignment:

Critique classmate's literature review on his or her chosen disease

Week 10

Spring Break

Week 11

Readings:

- Relevant chapter chosen by classmate
- Foxman, Molecular Tools, Chapters 7–10 (Omics analyses in Molecular Epidemiologic Studies, Determining the Reliability and Validity and Interpretation of a Measure in the Study Populations, Designing and Implementing a Molecular Epidemiologic Study, Study Conduct)

Assignment:

Critique classmate's literature review on his or her chosen disease

Week 12

Readings:

- Relevant chapter chosen by classmate
- Foxman, Molecular Tools, Chapters 11 and 12 (Think about Data Analysis When Planning a Study, Human and Animal Subject Protection, Biorepositories, Biosafety Considerations and Professional Ethics)

Assignment:

Critique classmate's literature review on his or her chosen disease

Week 13

Readings:

- Relevant chapter chosen by classmate
- Foxman, Molecular Tools, Chapter 13 (Future Directions)

Assignment:

Critique classmate's literature review on his or her chosen disease

Week 14

Final projects due

Week 15

Present final project to class

Critique and comment on classmates' final projects

Reproduced with the kind permission of C. A. Molgaard

examining the strengths and weaknesses of the student's proposal and provided an orientation to focused grantsmanship in neuroepidemiology after in-depth reading and discussion in the field. The course 'Introduction to Epidemiology' was a prerequisite for the class.

Summary

In this paper, we have defined the field of neuroepidemiology, presented several examples of research paradigms used in this subfield of epidemiology, presented concepts from developmental neurotoxicology of relevance to teaching in this field, offered suggestions for their integration into the teaching protocol used for a course in neuroepidemiology, and discussed a real-world digital learning neuroepidemiology class. The main epidemiologic emphasis for the teacher is the concept of programming as related to critical windows of exposure and vulnerability for the developing nervous system.

Within neuroepidemiology, the standard research paradigm in the past has been one of focusing on neurodegenerative diseases as an adjunct of the normal ageing process. Seminal research during the 1950s and 1960s introduced the notion of slow viruses with extremely long latency periods between exposure and onset of neurodegenerative disease. This notion was extended to slow toxins, for example, in the case of Western Pacific amyotrophic lateral sclerosis and the use of the cycad nut as a source of food and medicine among the Chamorro population on the island of Guam: although this hypothesis is controversial, the cycad nut is believed to contain several neurotoxins responsible for this endemic pattern of disease on Guam.

Recent research in developmental neurotoxicity has further extended this concept of extremely long-term (often in utero) exposure and onset of neurodegenerative disease much later in adult life. Diseases such as Parkinson's disease, lead toxicity, methyl mercury toxicity, schizophrenia, dyslexia, epilepsy, and autism, as well as many others, are thought to be examples of this process. Two explanatory theories are posited. In the first, the development of a specific function in an organism normally occurs late, and the display of pathologic function is not obvious until that point in development is reached. In the second, developmental structural or functional damage is masked by neural plasticity, so that effects are transient until much later during the lifespan (Rice and Barone 2000).

The interaction of the processes of ageing with neurotoxic exposures is now thought to be a common event. In normal ageing, the brain loses cells in some regions, as well as suffering declines in neurotransmitters and repair mechanisms. Neurotoxic exposures of long latency can accelerate such events and

increase functional disability. As populations in industrialized countries continue to age in the US, Canada, Japan, and Europe, the effect of such interaction may have significant economic and social impact on already overburdened public health systems. Research and teaching agendas in neuroepidemiology should also reflect these aspects of developmental neurotoxicity (see Box 19.2).

References

Adams, J., Barone, S. Jr, LaMantia, A., Philen, R., Rice, D. C., Spear, L., and Susser, E. (2000) Workshop to identify critical windows of exposure for children's health: Neurobehavioral Work Group summary. *Environmental Health Perspectives*, **108** Suppl. 3: 535–44.

Anderson, D. W. and Schoenberg, D. G. (1991) *Neuroepidemiology: A Tribute to Bruce Schoenberg*. Boca Raton, FL: CRC Press.

Andersson, M., Takkouche, B., Egli, I., Allen, H. E., and de Benoist, B. (2005) Current global iodine status and progress over the last decade towards the elimination of iodine deficiency. *Bulletin of the World Health Organization*, **83**: 518–25.

Batchelor, T. and Cudkowicz, M. E. (1999) *Principles of Neuroepidemiology*. Boston, MA: Butterworth-Heinemann.

Bigal, M., Serrano, D., Reed, M., and Lipton, R. (2008) Chronic migraine in the population: burden, diagnosis and satisfaction with treatment. *Neurology*, **71**: 559–66.

Foxman, B. (2012) *Molecular Tools and Infectious Disease Epidemiology*. New York: Elsevier/Academic Press.

Gilfillan, S. C. (1965) Lead poisoning and the fall of Rome. *Journal of Occupational Medicine*, **7**: 53–60.

Gluckman, P. and Hanson, M. (2005) *The Fetal Matrix: Evolution, Development and Disease*. Cambridge: Cambridge University Press.

Goldman, L. R. and Kodura, S. (2000) Chemicals in the environment and developmental toxicity to children: a public health and policy perspective. *Environmental Health Perspectives*, **108** Suppl. 3: 443–50.

Gorelick, P. B. and Alter, M. (1994) *Handbook of Neuroepidemiology*. New York: M. Dekker.

Guillette, E. A. (2000) Examining childhood development in contaminated urban settings. *Environmental Health Perspectives*, **108** Suppl. 3: 389–93.

Hallas, J., Bjerrum, L., Stovring, H., and Anderson, M. (2008) Use of a prescribed ephedrine/caffeine combination and the risk of serious cardiovascular events: a registry-based case-crossover study. *American Journal of Epidemiology*, **168**: 966–73.

Hernberg, S. (2000) Lead poisoning in a historical perspective. *American Journal of Industrial Medicine*, **38**: 244–54.

Hernstein, R. J.and Murray, C. (1996) *Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hough, R. L., Kolody, B., and Du Bois, B. (1993) 'The epidemiology of Alzheimer's disease and dementia among Hispanic Americans', in **C. A. Molgaard**, ed, *Neuroepidemiology: Theory and Method*. San Diego, CA: Academic Press, pp. 352–65.

Kurland, L. T. (1978) Geographic isolates: their role in neuroepidemiology. *Advances in Neurology*, **19**: 69–81.

Kurland, L. T. and Molgaard, C. A. (1981) The patient record in epidemiology. *Scientific American*, **245**: 54–63.

Kurtzke, J. F. (2000) Multiple sclerosis in time and space-geographic clues to cause. *Journal of Neurovirology*, **6**: 134–40.

Kurtzke, J. F. and Hyllested, K. (1979) Multiple sclerosis in the Faroe Islands. I. Clinical and epidemiologic features. *Annals of Neurology*, **5**: 6–21.

Kurtzke, J. F., Hyllested, K., and Heltberg, A. (1993) 'Multiple sclerosis in the Faroe Islands', in **C. A. Molgaard**, ed., *Neuroepidemiology: Theory and Method*. San Diego, CA: Academic Press, pp. 24–50.

Lewis, J. (1985) Lead poisoning: a historical perspective. *Environmental Protection Agency Journal*, **11**: 15.

Magna Publications. (2013) *Distance Education Report: April 1, 2013*. <http://www.magnapubs.com/newsletter/distance-education-report/issue/1440/>, accessed 5 November 2014.

Molgaard, C.A. (2002) Cognitive and social correlates of highly prevalent endemic diseases: the Rodney Dangerfield effect. *Headache*, **42**(2): 1–3.

Molgaard, C. A., ed. (1993) *Neuroepidemiology: Theory and Method*. San Diego, CA: Academic Press.

Molgaard, C. A. and Golbeck, A. L. (1992) Mad cows and Englishmen: bovine spongiform encephalopathy (BSE). *Neuroepidemiology*, **9**: 285–6.

Nelson, L. M., Tanner, C. M., Van Den Eeden, S. K., and McGuire, V. M., eds. (2004) *Neuroepidemiology: from Principles to Practice*. New York: Oxford University Press.

Olney, J. W., Farber, N. B., Wozniak, D. F., Jevtovic-Todorovic, V., and Ikonomidou, C. (2000) Environmental agents that have the potential to trigger massive apoptotic neurodegeneration in the developing brain. *Environmental Health Perspectives*, **108** Suppl. **3**: 383–8.

Oshinsky, D. M. (2005) *Polio: An American Story*. Oxford: Oxford University Press.

Osmond, C. and Barker, D. J. P. (2000) Fetal, infant, and childhood growth are predictors of coronary heart disease, diabetes, and hypertension in adult men and women. *Environmental Health Perspectives*, **108** Suppl. 3: 545–53.

Paschal, A. M., Ablah, E., Wetta-Hall, R., Molgaard, C. A., and Liow, K. (2005) Stigma and safe havens: a medical sociological perspective on African-American female epilepsy patients. *Epilepsy and Behavior*, **7**: 106–15.

Paschal, A. M., Hawley, S. R., and Sly, J. (2007) 'Epilepsy and community-based participatory research: some issues of social capital and awareness', in **F. Columbus**, ed., *New Research on Epilepsy and Behavior*. Happauge, NY: Nova Scotia Publishers, Inc., pp. 185–209.

Prince, M. (1998) Is chronic low-level lead exposure in early life an etiologic factor in Alzheimer's disease? *Epidemiology*, **9**: 618–21.

Rice, D. and Barone, S. Jr. (2000) Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. *Environmental Health Perspectives*, **108** Suppl. 3: 511–33.

Rothrock, J. F., Lyden, P. D., and Brody, M. L. (1993) 'The utility of stroke data banks in the epidemiology of cerebrovascular diseases', in **C. A. Molgaard**, ed., *Neuroepidemiology: Theory and Method*. San Diego, CA: Academic Press, pp. 287–305.

Rothrock, J. F., Lyden, P. D., Brody, M. L., Taft-Alvarez, B., Kelly, N., Mayer, J., and Wiederholt, W. C. (1993) An analysis of ischemic stroke in an urban southern California population: the UCSD Stroke Data Bank. *Archives of Internal Medicine*, **153**: 619–24.

Sarbo, K. (2012) *True Strength*. Cambridge, Mass: Da Capo Press.

Schulte, P. A. and Perera, F. P. (1993) *Molecular Epidemiology Principles and Practices*. San Diego, CA: Academic Press.

Silberstein, S. et al. (2005) The International Classification of Headache Disorders, 2nd edition (ICHD-II)—revision of criteria for 8.2 *Medication-overuse headache*. *Cephalalgia*, **25**: 460–5.

Tanner, C. M. and Ross, G. W. (2004) 'Neuroepidemiology: fundamental considerations', in **L. M. Nelson**, **C. M. Tanner**, **S. K. Van Den Eeden**, and **V. M. McGuire**, eds, *Neuroepidemiology: From Principles to Practice*. Oxford: Oxford University Press, pp. 1–22.

Tilson, H. A. (2000) New horizons: future directions in neurotoxicology. *Environmental Health Perspectives*, **108 Suppl. 3**: 439–42.

Weiss, B. and Landrigan, P. J. (2000) The developing brain and the environment: an introduction. *Environmental Health Perspectives*, **108** Suppl. 3: 373–4.

Zweifler, R., Mendizabal, J., Shaw, A., Cunningham, S., and Rothrock, J. F. (2002) Hospitalization presentation after stroke in a community sample: the Mobile Stroke Project. *Southern Medical Journal*, **95**: 1263–8.

Chapter 20

Reproductive epidemiology

Jørn Olsen and Ellen Aagaard Nøhr

Introduction to reproductive epidemiology

This chapter deals with teaching the epidemiology of reproductive health (or parts of reproductive health) to students who already know something about basic epidemiology. The aim is to provide an insight into methods and pitfalls when doing epidemiologic studies on reproductive outcomes. It is not a course on what is known in reproductive health about determinants or biological mechanisms. Furthermore, it only covers part of reproductive health. According to the World Health Organization (WHO) (from the 'Cairo Conference' in 1994, <http://www.iisd.ca/cairo.html>), reproductive health is

> a state of complete physical, mental and social wellbeing, and not merely the absence of disease or infirmity, in all matters relating to the reproductive system and its functions and processes. Reproductive health, therefore, implies that people are able to have a satisfying and safe sex life and that they have the capacity to reproduce and the freedom to decide if, when, and how often to do so.

The WHO includes diseases, well-being, and perhaps even happiness as a potential health outcome. We do not intend to cover such a broad area but will mainly focus upon how diseases or organ functions are studied in epidemiology. We exclude, however, sexually transmitted diseases and a number of other diseases (like cancer of the cervix) because they would usually be discussed in a course on infectious disease epidemiology or a course in cancer epidemiology.

Exposures or conditions during the early time period of life, perhaps especially the time from conception to birth, probably plays an important role for understanding a number of diseases with an onset much later in life. We think that this topic should be included in a course on reproductive epidemiology if a full seminar in life course epidemiology is not provided.

In this chapter, we will not repeat definitions or terminology to any great length but refer the reader to Nguyen and Wilcox's papers (2005a, b), the work by Kline et al. (1989), or the official International Epidemiology Association Dictionary (Porta et al. 2008).

We provide a set of highly selected references to the teacher and to some extent also to the students. We strongly suggest that these references are read and understood by the teacher before starting the course. We also refer to the newest textbooks in reproductive epidemiology and to two comprehensive textbooks in epidemiology with chapters on reproductive health (Savitz et al. 2002; Ahrens and Pigeot 2005; Rothman et al. 2008; Wilcox 2010; Louis and Platt 2011).

Teaching objectives

The aim of this course should be to make the students aware of the conditions and problems that are common and specific for the different parts of reproductive epidemiology. At the end of the course, the student should be able to read the epidemiologic literature on reproductive outcomes with critical eyes. S/he should be aware of our research options and their limitations. For a student with a good background in epidemiologic methods, the course should make it possible for him/her to write sound protocols in reproductive health.

It is important that the course address what is known and what is not known on the topics the teacher addresses and that the teacher is willing to share his/her uncertainties and limitations with the students. It is our experience that students are generous and willing to accept an inexperienced teacher if s/he is open about this.

Teaching content/syllabus

The syllabus

We suggest a 20-hour course structured as ten 2-hour sessions (Table 20.1), including lectures and class sessions in which the students present and discuss the results of their group work. If less time is available, we suggest having only sessions 1, 2, and 9.

Session 1: reproductive health in a global perspective

The first lecture is a descriptive overview of the key outcomes in reproductive health, including data from both affluent and low income countries. It may touch upon the Millennium Developmental Goals (see <http://www.un.org/milleniumgoals/>; White et al. 2006; The Lancet 2013) that did not include reproductive health. The disparity between rich and poor countries with respect to maternal and neonatal morbidity and mortality should be made clear to the students, and the teacher may note that most of our research takes place in affluent countries, where health problems are much less frequent than in low-income

Table 20.1 Reproductive epidemiology course structure

Session	Content
1	Reproductive health in a global perspective
2	What makes reproductive epidemiology special? Major avoidable risk factors
3	Fertility
4	Foetal death and congenital malformations
5	Foetal growth
6	Pregnancy complications
7	Sexual health
8	Health in a life course perspective 1—childhood diseases and development
9	Health in a life course perspective 2—adult diseases
10	Final exam

parts of the world. The willingness to invest in reproductive health research in low-income countries has to some extent been driven by the desire to control global population growth.

Methods used to collect data to monitor reproductive health in local and international settings and to do descriptive studies should be presented. Many countries have kept birth registration systems for decades, and routine registration of congenital malformations is also done in some countries or regions. Most routine health surveys will include some data on reproductive health (Glasier et al. 2006).

Session 2: what makes reproductive epidemiology special? Major avoidable risk factors for adverse pregnancy outcomes

Building upon the students' skills in epidemiology, the teachers should focus upon what is different and unique in reproductive health epidemiology. The overview can be based upon the works by Bracken (1984), Olsen and Basso (2005), Wilcox (2010), and Louis and Platt (2011) and should provide an introduction to the overall methodological challenges when studying reproductive outcomes. The teaching should at least address the following aspects. The unit of observation may involve not only the mother but also the partner and the child. The time period of observation may be short—nine months or less—or it may be very long, as in life course studies. The outcome we study may be repeated if the couple has more than one child. There is a tendency to repeat reproductive failures, and past experience may impact changes in exposure and

thus generate different confounder and collider structures that are not trivial (Basso and Wilcox 2011; Wilcox et al. 2011; Howards et al. 2012). The time period of pregnancy is under intensive medical surveillance in most countries and we can only study the associations that remain between exposures and effects after antenatal care has done its part. Parents are usually highly motivated to take part in research because they care for the health of the unborn child. This trust is not to be misused. Results are to be communicated with care and caution. Exaggerating risks may cause unnecessary stress, grief, and guilt.

This session should also present methods to study major avoidable risk factors, given that the overall aim of reproductive epidemiology is to improve health in mothers and children. These factors could be divided into those that are modifiable on an individual level, such as diet, medicine, and occupational stress loads, and those that are modifiable on the community level, such as contamination of drinking water, and some types of pollution. In reproductive health, the timing of exposure may be more important than in other studies and we may sometimes see a dose-effect response rather than a dose-response effect: heavy exposures could lead to infertility or foetal loss, and insufficient exposure could lead to a subtle change in organ functions.

When studying an exposure that clusters with many other exposures (e.g. dietary factors), an observational study has considerable limitations, but it may be possible to do randomized trials that manipulate a single exposure like a vitamin, *n*-3 fatty acids, milk, etc. The students should be aware of the opportunities and limitations these studies have in pregnant women; in particular, ethical aspects should be addressed. The time span from exposure to effect is often short and thus provides an excellent opportunity for doing a study with high compliance. The literature contains many excellent examples of randomized trials that have produced data that can be used for improving reproductive health (e.g. see the WHO Reproductive Health Library (<http://www.who.int/hrp/rhl/en/>)). Some good examples of methods that have tried to overcome some of the ethical issues could be presented, such as trials aiming at altering lifestyle factors in already exposed women (Bech et al. 2007), and trials on vitamin supplementation in developing countries.

Some features solely related to reproduction facilitate special observational designs. The presence of several pregnancies related to the same woman but with different levels of the same exposure allows us to disentangle the influence from social confounders by using sibling designs (Olsen et al. 1997; Lambe et al. 2006; Obel et al. 2011). Since some genes are randomly allocated at the time of conception, studies can use a case-parent design to examine genetic and environmental determinants of disease (Wilcox et al. 1998; Wilcox 2010). Alternatively, the lecture could take up the Mendelian randomization design, which

may be used when there is a strong link between a genotype and a phenotype that is suspected to cause a disease (Smith and Ebrahim 2004). The genotype will then be an 'instrumental variable' for this exposure and will normally not be expected to be seriously confounded by lifestyle factors because genes are allocated at random at the time of conception. This design has been used to corroborate the possible causal link between folic acid intake and neural tube defects (Sheehan et al. 2008). It has also been used to study potential effects of mobile phone exposures (Sudan et al. 2013).

Session 3: fertility

Teaching should address the problems we face when studying factors that may impact the biological capacity to conceive, called fecundity or fecundability (the probability of becoming pregnant in a given menstrual cycle). This is a variable that is not directly observable but has been estimated by using data on fertility, time to pregnancy, or specific markers like semen quality or menstrual disorders.

Fertility is defined by demographers as the number of live births women have from age fifteen to age forty-five divided by the total number of women at that age. At the population level, the fertility ratio will only produce a reliable estimate of population growth and family sizes under steady state situations.

Since subfecundity involves both partners, a measure like time to pregnancy (TTP) may be used. The first studies using this measure date back to the 1980s (Rachootin and Olsen 1982). We now know much more about the limitations we face when using this tool (Bonde et al. 2006).

More indirect ways to measure male or female fecundity involve studies on twinning, since dizygotic twins may represent high fecundity. Biomarkers like semen quality, menstrual disorders, and diseases of the ovary have also been used.

The session should also address the issue of monitoring fecundity over time. The paper on decline in semen quality (Carlsen et al. 1992) provides a good discussion on limitations and sources of bias in ecological studies. The paper by Sallmen et al. (2005) can be used to illustrate the difficulties in using TTP measures to monitor past changes in fecundity. It may also be worthwhile to spend a little time on a measure called 'the standardized fertility ratio', which has been used to estimate changes in fecundity before or after a given exposure (Levine et al. 1981).

Session 4: foetal death and congenital malformations

This lecture should present data on the change in rate of foetal death over gestational time. Foetal loss may occur early before the pregnancy is recognized.

Exposures that move an abortion forward in time may move it into the undetected time period and thus make the exposure appear to be protective.

Few studies have used biomarkers to detect pregnancies before clinical diagnosis, since these studies are complicated and expensive to do (Bonde et al. 1998), but they have shown that about two-thirds of all abortions happen before the pregnancy is clinically diagnosed (Wilcox 2010). Early abortions may well have a different etiology than later abortions, as most early abortions have chromosomal abnormalities. The gestational cut-off level that divides an abortion from a stillbirth differs between different countries but a cut-off between twenty to twenty-four weeks is frequently used. Problems related to estimating gestational age may well affect monitoring of abortion and stillbirth rates.

Since the timing of foetal death often differs from the time of expulsion, from days to several weeks, reverse causation is a possible source of bias. If foetal death induces exposures, like lifestyle factors (e.g. coffee or alcohol use) or medical treatment (like use of painkillers), reverse causation may explain positive associations between the exposure and the abortion rate in a follow-up study.

Students should also be aware of the fact that induced abortions may impact abortion rates if the timing and frequency of these abortions are not known (Olsen 1984).

Studies of foetal death will include many different pathologies and may therefore be too unspecific to have sufficient power to detect exposures that only affect deaths related to a specific cause. Subclassification by gestational age may help but may not be enough. Access to the aborted tissue could provide more detailed diagnostic data but such data are difficult to obtain. Medical records on stillbirths may provide information to define causes of death, and several classification systems have been suggested.

The use of directed acyclic graphs has clarified that previous methods of analysing data on reproductive outcomes are often misleading. The problem is mainly related to treating intermediate variables as if they are confounders (Howards et al. 2007; Basso 2011; Howards et al. 2012). Adding intermediates to the model without taking causes of the intermediating variable into consideration may cause confounding not present before.

Congenital malformation

Congenital malformations or abnormalities are by definition present at birth, although many are not diagnosed until later in life and some may remain undiagnosed. It is important for students to understand that the proportion of malformation at birth is a prevalent measure, often measured at the time of birth or at 1 year of age. An estimate of the cumulative incidence (risk) would require

follow-up of all conceptions, with complete registration of all malformations during this time period (and without competing risks). An exposure that may have no etiologic role but impacts the probability of survival of an affected foetus will be associated with the prevalence of congenital malformations at birth and will present itself as a potential protective factor—a bit like early suicide will prevent later cancers—illustrating that prevention is not always better than cure.

Congenital malformations are sometimes difficult to diagnose and many studies exclude so-called minor malformations, like dislocation of the hip, undescended testis, and birth marks, because the diagnostic validity for these diagnostic categories may be too poor.

Students should know that large registers on malformations exist in many countries, mainly because of the thalidomide disaster in the 1960s (McBride 1961). Since this finding, we have seen no similar strong links between use of medicine during pregnancy and congenital malformations, and the risk of this happening again is small but still present. For ethical reasons, new drugs are not tested on pregnant women before they come onto the market, and the evidence we have on the potential foetotoxic effects come mainly from animal studies. We still have very few human studies on functional defects, especially long-term effects.

Students should also be aware of the importance of prenatal screening for congenital malformations which may lead to termination of pregnancy. Unless we have access to valid screening data, we will only be able to identify associations that remain after abortions have been induced for clinical reasons. Frequent use of ultrasound examination during pregnancy will provide new important measures such as foetal growth and malformations diagnosed in utero but it will also produce new sources of bias.

Diagnosing congenital malformations is not based on etiologic principles which take the time of organ development into consideration, such as using malformations related to the cranial neural crest and the cells that migrate from this structure (Zierler and Rothman 1985; Ingstrup et al. 2013).

Standardization of diagnosing congenital malformations and data collection are done at the international level by organizations like EUROCAT (<http://www.eurocat-network.eu>) or the International Clearinghouse for Birth Defects (<http://www.icbdsr.org>).

Session 5: foetal growth

Since birth weight (BW) is routinely recorded in most health-care systems, numerous studies have addressed the associations between BW and external exposures; such studies are often justified by the fact that low BW is strongly

correlated with infant morbidity and mortality. However, this activity reflects the availability research options more than the importance of the research. The teaching should underscore that BW is a function of foetal growth and the duration of pregnancy and that it has causes that may be more important than BW in itself. Measures like small for gestational age (SGA) or *Z*-scores try to take gestational age into consideration but may make good data on BW poor by including a notorious frequently misclassified measure on gestational age at birth, especially if historical data are used. Usually, the interest in public health is to identify deviations from the genetically predicted BW and the attained BW. The teaching may include how this has been done with some success by using family histories of BWs (Skjaerven et al. 2000). Low BW and preterm birth often reflect underlying health problems and may be a marker rather than a cause in itself (Basso and Wilcox 2011).

Being SGA means being smaller than expected at a certain period in time, and the comparison should as usual be the population that have reached the same gestational age, born or not born (Marsal et al. 1996). Often, an SGA measure is based on the BW distribution among those who were born in that gestational week, and early in gestation many of these are small due to health problems.

Many studies still dichotomize BWs at 2,500 grams, a practice which should be avoided (Wilcox 2001). This level was set to capture preterm births in countries where data on gestational age often were missing, since most (but not all) of babies born with a BW of less than 2,500 grams are preterm.

The body mass index measure at birth (the ponderal index; Cole et al. 1997) should be presented, and it should be stressed that the only reason for dividing weight with height raised to the power of 3 (rather than 2 as in the BMI) was to obtain a more symmetrical distribution.

Wilcox (2001) has advocated that it is not the actual BW distribution and its central tendency we should be concerned about but the fraction of outliers from the Gaussian distribution (i.e. residuals). Basso et al. (2006) provided a simulation study to show that a single serious pathology associated with foetal growth and mortality could generate the BW distributions that we see at the population level.

Organ growth can to some extent be estimated at birth by measuring head circumference (brain growth), abdominal circumference (liver), or placental weight. Organ growth or interference with organ growth may be of importance for later organ function and disease susceptibility (Barker 1995).

Examples of historical studies of interest to examine in detail are Yerushalmy's (1972) study on smoking and BW. He illustrated that it may be more powerful to study what will be incompatible with the hypothesis than to set up

another replication study (falsification vs verification). He believed that babies born to smoking mothers were smaller because of the constitution of the mothers (a smoker's effect) rather than the exposure itself (a smoking effect). He therefore predicted that women who smoked in one pregnancy but not in another would also get small children in the non-smoking pregnancy. Although his hypothesis was wrong, his approach to put a hypothesis to a critical test is useful and worth considering in many other situations. For example, women with diabetes more often give birth to children with congenital malformations; therefore, these malformations could theoretically be caused by the same genetic factors that cause diabetes. However, if this were the case, the association should also be present in children born to fathers with diabetes (which was not found; Wu et al. 2012).

Session 6: pregnancy complications

This session should focus upon the diseases and complications that are specific to a pregnancy, such as pre-eclampsia, gestational diabetes, abruptio placentae, and placenta praevia. Although many diseases are modified by pregnancy, that topic can be addressed in regular disease-oriented lectures.

The definitions of these complications will in most cases depend on the clinical diagnosis, and it is important to present to the students the implications of sensitivity, specificity, and disease frequency for interpreting the results of epidemiologic findings. For example, specificity of recording often needs to be high when studying etiology of a certain complication to avoid including non-cases in the case group, while a high sensitivity is preferable if the case definition is used to exclude unwanted co-morbidity.

The timing of the detection of many of these diseases will depend upon the frequency and quality of the antenatal care programme. When hypertension or proteinuria will be detected will also depend upon how often screening is done. In addition, selective screening of high-risk groups (e.g. screening for gestational diabetes in overweight women) may be carried out, and this may influence internal comparisons.

By taking into account the onset of pregnancy complications, the gestational-age-specific incidence can be estimated instead of the cumulative incidence over an entire pregnancy, thus allowing an examination of time-dependent effects. For example, gestational diabetes may have a much larger effect on foetal growth if onset occurs in week 20 rather than in late pregnancy.

Preterm birth is probably the main pregnancy complication (Behrman and Butler 2007). Students should know that preterm birth is solely defined by gestational time and refers to a birth before thirty-seven weeks of gestation. A premature birth is a clinical condition that requires data to indicate that the child

in fact is premature. Preterm birth, especially very preterm birth (born before thirty-two weeks of gestation), is strongly linked to severe and long-term health consequences or death.

An accurate diagnosis of preterm birth requires data on the time of conception. Since this is usually only known precisely in pregnancies where assisted reproductive technologies have been used, the expected date of the delivery is normally based upon the last menstrual period method (LMP) by applying Naegele's rule (take the first day of LMP, subtract 3 months and add 1 week; Nägele 1836) or estimates of early foetal size by the use of ultrasound. Both methods are subject to bias: the LMP period relies on a regular menstrual cycle and accurate recall, while ultrasound is based upon the assumption of a similar growth pattern for all foetuses in early pregnancy, an assumption that is probably not valid but good enough for making clinical prediction with a sufficient precision. This justification is more questionable when studying the effect of exposures that may impact early foetal growth (Henriksen et al. 1995).

Preterm births are frequent (from 4–20%), and in spite of huge investments, relatively little is known on how to prevent them. Preterm births, like most other reproductive failures, have a tendency to repeat themselves. The reason for that could be that they have causes that do not change from one pregnancy to the next, like genetic factors or time-stable external causes.

Pre-eclampsia may be used to illustrate a large number of changing hypotheses over time, often inspired by its interesting epidemiology (Wilcox 2010). Still, we are far from being able to prevent a larger fraction of pre-eclampsia.

Delivery complications include abnormal foetal/placental presentation, prolonged labour (dystocia), foetopelvic disproportion, instrumental delivery, caesarean section, and haemorrhage. Studying risk factors for these endpoints is heavily vulnerable to confounding by indication. Clinicians make decisions that are often based upon information that cannot be read by a machine and therefore cannot be adjusted in a statistical analysis. Randomized trials may be the only option in many situations, and students should be familiar with Cochrane reviews on trials in reproductive health (see <http://www.cochrane.org>).

Session 7: sexual health

The role of epidemiology in studying sexual health has been limited, at least for the broader picture as described in WHO's definition of reproductive health. However, epidemiologists take part in getting descriptive statistics on sexual health. A number of youth surveys are being conducted at regular intervals in order to get data on sexual debut, sexually transmitted diseases, use of contraceptives, age of menarche, and sexual practices. These data are not only of

interest for planners of health education but can also be of importance for setting up more targeted studies on potential risk factors for sexual health problems.

Session 8: health in a life course perspective 1—childhood diseases and development

It is not unexpected that the prenatal time period has an impact on early childhood development and the diseases that occur in early life. Students should be familiar with the Apgar score, which is routinely used in many countries as the first measure of child health. They should be familiar with cognitive and motor developmental milestones and how to measure them (Harrington et al. 2007). They should be aware of the method problems entailed in moving from the conception/pregnancy time axis to the birth time axis at different points in gestational time. Onset of some developmental milestones may, for example, be more related to the time of conception than the time of birth.

Prenatal exposures may in principle affect all organ systems and thus susceptibility to a number of diseases that occur early in childhood, perhaps modified by the growth trajectories after birth. Teaching examples may use examples of long-term consequences of foetotoxic exposures like alcohol, lead, or mercury and include the evidence we have for a prenatal set of causes that activate the first step in, for example, the etiology of childhood leukaemia or cancer of the testis.

Session 9: health in a life course perspective 2—adult diseases

Most of the new research activities in epidemiology address the possible importance of the early origin of adult diseases. Many believe that foetal programming shapes susceptibility for diseases that may manifest themselves much later in life; unfortunately, we still have limited data to study this.

The students should be informed about the important work done by Barker and Forsdal, who showed that time periods of food shortage may induce insulin resistance in the foetus and that this may then lead to an increased risk of cardiovascular disease in adult life as well as diabetes and obesity (Barker 1995; Gluckman and Hanson 2004). It is now believed that many other factors such as stress and infection may also have a programming effect on the foetus (Brown et al. 2005; Li et al. 2011). Similarly, exposure to toxins like alcohol, mercury, pesticides, and perfluorocarbons may also cause permanent health problems that need not be detectable at birth. The challenge is to set up studies that can address these issues, and the students should be introduced to the many ongoing cohort studies (e.g. see <http://www.birthcohorts.net>) and some of the most important animal studies that support the idea of foetal programming. Epigenetic

changes may be a topic to address in this session. Problems in analysing life course data should be addressed. The first and most important problem is to know which data to collect (at present we just collect as much data as possible). The next problem is related to the statistical analyses that of course have to be driven by sound hypotheses (De Stavola et al. 2006).

The students should also be introduced to the classical study on two generations effects done by Herbst et al. (1971). They showed that women who during pregnancy took the oestrogen-like hormone diethylstilboestrol to prevent a miscarriage or preterm birth had girls with an increased risk of vaginal cancer. This example may introduce a discussion on similar possible etiologies for breast cancer, testis cancer, and certain childhood cancers, especially leukaemia. Long-term follow-up is often complicated and requires different data sources and search strategies (Klebanoff et al. 1998). More recent studies indicate that lifestyle factors like smoking may impact fecundity in the next generation (Ramlau-Hansen et al. 2007).

Teaching method and format

We believe in active participation from students in all parts of the learning process. Two textbooks in reproductive epidemiology, published since 2010, could comprise important course material, but a number of structured lectures are also needed (Wilcox 2010; Louis and Platt 2011). A combination of these lectures and discussions of published papers and exercises usually work well. We suggest that the two-hour sessions allow for at least one hour of problem solving.

If possible, the group work can be centred around analysing an existing data source on reproductive failures, but that will require that the students are familiar with a statistical software package and that the questions they need to address are simple and very specific. If that not the case, the sessions end up being about how to use a particular software package. Most concepts are best dealt with as epistemological problems that need to be discussed and challenged.

Assessing students' achievements

We believe it is important to measure students' achievements after a course, whether this has to be used to provide a grade or not. It is necessary for the teacher to see how well the objective was achieved and for the students to get an idea of their own performance.

We are not particularly fond of multiple-choice questions because epidemiologic questions seldom have just one answer. Most of these exercises end up with students guessing what they think the professor thinks.

Rather, the students should be aware of what they know and what they do not know, and it is also important that they know what the teacher does not know and at best what nobody knows. A good exam allows for uncertainty and the fact that questions may have more than one answer.

An obvious exam question is to let students discuss a recent publication within one of the areas that was covered in the lectures. Such a review could be structured as a regular peer review by letting the students comment on the clarity of the aims of the study, whether the method and data were appropriate for the aims, whether the analyses were informative and sufficient, whether the authors addressed weak parts of the study in a critical way, and whether the authors presented a conclusion that was justified by the analyses. If time permits, the students could be asked to suggest alternative designs to address the aims of the study. By doing this, the exam allows for open and flexible suggestions that will illustrate whether the student is able to prioritize between very important and less important method issues. Still, carefully designed problem-solving exercises should also be used.

References

Ahrens, W. and Pigeot, I., eds. (2005) *Handbook of Epidemiology*. Berlin: Springer.

Barker, D. J. (1995) Fetal origins of coronary heart disease. *British Medical Journal*, **311**: 171–11.

Basso, O. and Wilcox, A. J. (2011) Might rare factors account for most of the mortality of preterm babies? *Epidemiology*, **22**: 320–7.

Basso, O., Wilcox, A. J., and Weinberg, C. R. (2006) Birthweight and mortality: causality or confounding? *American Journal of Epidemiology*, **164**: 303–11.

Bech, B. H., Obel, C., Henriksen, T. B., and Olsen, J. (2007) Effect of reducing caffeine intake on birthweight and length of gestation: randomised controlled trial. *British Medical Journal*, **334**: 409–15.

Behrman, R. E. and Butler, A. S., eds. (2007) *Preterm Birth: Causes, Consequences, and Prevention*. Washington, DC: National Academies Press.

Bonde, J. P., Hjollund, N. H., Jensen, T. K., Ernst, E., Kolstad, H., Henriksen, T. B., Giwercman, A., Skakkebaek, N. E., Andersson, A. M., and Olsen, J. (1998) A follow-up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: design and methods. *Reproductive Toxicology*, **12**: 19–27.

Bonde, J. P., Joffe, M., Sallmén, M., Kristensen, P., Olsen, J., Roeleveld, N., and Wilcox, A. (2006) Validity issues relating to time-to-pregnancy studies of fertility. *Epidemiology*, **17**: 347–9.

Bracken, M. B., ed. (1984) *Perinatal Epidemiology*. New York: Oxford University Press.

Brown, A. S., Schaefer, C. A., Quesenberry, C. P. Jr, Liu, L., Babulas, V. P., and Susser, E. S. (2005) Maternal exposure to toxoplasmosis and risk of schizophrenia in adult offspring. *American Journal of Psychiatry*, **162**: 767–73.

Carlsen, E., Giwercman, A., Keiding, N., and Skakkebaek, N. E. (1992) Evidence for decreasing quality of semen during past 50 years. *British Medical Journal*, **305**: 609–13.

Cole, T. J., Henson, G. L., Tremble, J. M., and Colley, N. V. (1997) Birthweight for length: ponderal index, body mass index or Benn index? *Annals of Human Biology*, **4**: 289–98.

De Stavola, B. L., Nitsch, D., dos Santos Silva, I., McCormack, V., Hardy, R., Mann, V., Cole, T. J., Morton, S., and Leon, D. A. (2006) Statistical issues in life course epidemiology. *American Journal of Epidemiology*, **163**: 84–96.

Glasier, A., Gülmezoglu, A. M., Schmid, G. P., Moreno, C. G., and Van Look, P. F. (2006) Sexual and reproductive health: a matter of life and death. *Lancet*, **368**: 1595–607.

Gluckman, P. D. and Hanson, M. A. (2004) Living with the past: evolution, development, and patterns of disease. *Science*, **305**: 1733–6.

Harrington, D. J., Redman, C. W., Moulden, M., and Greenwood, C. E. (2007) The long-term outcome in surviving infants with Apgar zero at 10 minutes: a systematic review of the literature and hospital-based cohort. *American Journal of Obstetrics and Gynecology*, **196**: 463–5.

Henriksen, T. B., Wilcox, A. J., Hedegaard, M., and Secher, N. J. (1995) Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. *Epidemiology*, **6**: 533–7.

Herbst, A. L., Ulfelder, H., and Poskanzer, D. C. (1971) Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *New England Journal of Medicine*, **284**: 878–81.

Howards, P. P., Schisterman, E. F., and Heagerty, P. J. (2007) Potential confounding by exposure history and prior outcomes: an example from perinatal epidemiology. *Epidemiology*, **18**: 544–51.

Howards, P. P., Schisterman, E. F., Poole, C., Kaufman, J. S., and Weinberg, C. R. (2012) 'Toward a clearer definition of confounding' revisited with directed acyclic graphs. *American Journal of Epidemiology*, **176**: 506–11.

Ingstrup, K. G., Liang, H., Olsen, J., Nohr, E. A., Bech, B. H., Wu, C. S., Christensen, K., and Li, J. (2013) Maternal bereavement in the antenatal period and oral cleft in the offspring. *Human Reproduction*, **28**: 1092–9.

Klebanoff, M. A., Zemel, B. S., Buka, S., and Zierler, S. (1998) Long-term follow-up of participants in the Collaborative Perinatal Project: tracking the next generation. *Paediatric and Perinatal Epidemiology*, **12**: 334–46.

Kline, J., Stein, Z., and Susser, M. (1989) *Conception to Birth: Epidemiology of Prenatal Development*. New York: Oxford University Press.

Lambe, M., Hultman, C., Torrång, A., MaCabe, J., and Cnattingius, S. (2006) Maternal smoking during pregnancy and school performance at age 15. *Epidemiology*, **17**: 524–30.

Levine, R. J., Symons, M. J., Balogh, S. A., Milby, T. H., and Whorton, M. D. (1981) A method for monitoring the fertility of workers. 2. Validation of the method among workers exposed to dibromochloropropane. *Journal of Occupational Medicine*, **23**: 183–8.

Li, J., Vestergaard, M., Obel, C., Cnattingus, S., Gissler, M., and Olsen, J. (2011) Cohort profile: The Nordic Perinatal Bereavement Cohort. *International Journal of Epidemiology*, **40**: 1161–7.

Louis, G. M. B. and Platt, R. W. (2011) *Reproductive and Perinatal Epidemiology*. New York: Oxford University Press.

Marsal, K., Persson, P. H., Larsen, T., Lilja, H., Selbing, A., and Sultan, B. (1996) Intrauterine growth curves based on ultrasonically estimated foetal weights. *Acta Paediatrica*, **1985**: 843–8.

McBride, W. G. (1961) Thalidomide and congenital malformations. *Lancet*, **2**: 1358.

Nguyen, R. H. and Wilcox, A. J. (2005a) Terms in reproductive and perinatal epidemiology: I. Reproductive terms. *Journal of Epidemiology and Community Health*, **59**: 916–9.

Nguyen, R. H. and Wilcox, A. J. (2005b) Terms in reproductive and perinatal epidemiology: 2. Perinatal terms. *Journal of Epidemiology and Community Health*, **59**: 1019–21.

Obel, C., Olsen, J., Henriksen, T. B., Rodriguez, A., Järvelin, M. R., Moilanen, I., Parner, E., Linnet, K. M., Taanila, A., Ebeling, H., Heiervang, E., and Gissler, M. (2011) Is maternal smoking during pregnancy a risk factor for hyperkinetic disorder? – Findings from a sibling design. *International Journal of Epidemiology*, **40**: 338–45.

Olsen, J. (1984) Calculating risk ratios for spontaneous abortions: the problem of induced abortions. *International Journal of Epidemiology*, **13**: 347–50.

Olsen, J. and Basso, O. (2005) 'Reproductive epidemiology', in **W. Ahrens** and **I. Pigeot**, eds, *Handbook of Epidemiology*. Berlin: Springer, pp. 1043–109.

Olsen, J., Schmidt, M. M., and Christensen, K. (1997) Evaluation of nature-nurture impact on reproductive health using half-siblings. *Epidemiology*, **8**: 6–11.

Porta, M., ed, Greenland, S., Last, J. M., and associate eds. (2008) *A Dictionary of Epidemiology. A Handbook Sponsored by the IEA* (5th edn). Oxford: Oxford University Press.

Rachootin, P. and Olsen, J. (1982) Prevalence and socioeconomic correlates of subfecundity and spontaneous abortion in Denmark. *International Journal of Epidemiology*, **11**: 245–9.

Ramlau-Hansen, C. H., Thulstrup, A. M., Storgaard, L., Toft, G., Olsen, J., and Bonde, J. P. (2007) Is prenatal exposure to tobacco smoking a cause of poor semen quality? A follow-up study. *American Journal of Epidemiology*, **165**: 1372–9.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008) *Modern Epidemiology* (3rd edn). Philadelphia, PA: Lippincott Williams & Wilkins.

Sallmen, M., Weinberg, C. R., Baird, D. D., Lindbohm, M. L., and Wilcox, A. J. (2005) Has human fertility declined over time? Why we may never know. *Epidemiology*, **16**: 494–9.

Savitz, D. A., Hertz-Picciotto, I., Poole, C., and Olshan, A. F. (2002) Epidemiologic measures of the course and outcome of pregnancy. *Epidemiologic Reviews*, **24**: 91–101.

Sheehan, N. A., Didelez, V., Burton, P. R., and Tobin, M. D. (2008) Mendelian randomisation and causal inference in observational epidemiology. *PLoS Medicine*, **5**: e177.

Skjaerven, R., Gjessing, H. K., and Bakketeig, L. S. (2000) New standards for birthweight by gestational age using family data. *American Journal of Obstetrics and Gynecology*, **183**: 689–96.

Smith, G. D. and Ebrahim, S. (2004) Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology*, **33**: 30–42.

Sudan, M., Kheifets, L., Arah, O. A., and Olsen, J. (2013) Cell phone exposures and hearing loss in children in the Danish National Birth Cohort. *Paediatric Perinatal Epidemiology*, **27**: 247–57.

The Lancet. (2013) Health and the post-2015 development agenda. *Lancet*, **381**: 699.

White, A. C., Merrick, T. W., and Yasbeck, A. S. (2006) *Reproductive Health: The Missing Millenium Development Goal: Poverty, Health and Development in a Changing World*. Washington, DC: The World Bank.

Wilcox, A. J. (2001) On the importance—and the unimportance—of birthweight. *International Journal of Epidemiology*, **30**: 1233–41.

Wilcox, A. J. (2010) *Fertility and Pregnancy: An Epidemiologic Perspective*. New York: Oxford University Press.

Wilcox, A. J., Weinberg, C. R., and Basso, O. (2011) On the pitfalls of adjusting for gestational age at birth. *American Journal of Epidemiology*, **174**: 1062–8.

Wilcox, A. J., Weinberg, C. R., and Lie, R. T. (1998) Distinguishing the effects of maternal and offspring genes through studies of 'case-parent triads'. *American Journal of Epidemiology*, **148**: 893–901.

Wu, C. S., Nohr, E. A., Bech, B. H., Vestergaard, M., and Olsen, J. (2012) Long-term health outcomes in children born to mothers with diabetes: a population-based cohort study. *PLoS One*, **7**: e36727.

Yerushalmy, J. (1972) Infants with low birthweight born before their mothers started to smoke cigarettes. *American Journal of Obstetrics and Gynecology*, **112**: 277–84.

Zierler, S. and Rothman, K. J. (1985) Congenital heart disease in relation to maternal use of Bendectin and other drugs in early pregnancy. *New England Journal of Medicine*, **313**: 347–52.

Chapter 21

Teaching chronic respiratory disease epidemiology

Josep M. Antó

Introduction to teaching chronic respiratory disease epidemiology

The development of respiratory disease epidemiology is relatively recent compared to other disease-oriented epidemiologies, such as those devoted to infectious disease, cancer, or cardiovascular disease. It should therefore not be surprising that respiratory disease epidemiology, with few exceptions, is not yet covered in most health teaching programmes. In the last decade, however, its importance has largely increased as a result of two important factors: the recognition that chronic airway diseases, mainly asthma and chronic obstructive pulmonary disease (COPD), impose a high burden of health to the society, and the sustained increase in epidemiological research to understand chronic respiratory diseases (Speizer and Tager 2007).

It is therefore predicable that more lectures and courses in respiratory epidemiology will be featured in postgraduate teaching programmes of both public health and respiratory medicine. This chapter aims to facilitate the establishment of introductory courses and sessions of respiratory epidemiology in a variety of teaching contexts. Teachers that do not have an expert knowledge on respiratory epidemiology may find helpful to use updated reviews and guidelines such as those by the Global Initiative for Asthma (GINA) and the Global Initiative for Chronic Obstructive Lung Disease (GOLD; Global Initiative for Asthma 2014; Global Initiative for Chronic Obstructive Lung Disease 2014).

This chapter primarily refers to chronic obstructive respiratory disease and asthma. Although this is a selective approach, other types of diseases, such as cancer and infections, will be covered elsewhere in this book. In wider programmes of respiratory epidemiology, attention could also be devoted to other diseases such as sleep apnoea, idiopathic pulmonary fibrosis, cystic fibrosis, or sarcoidosis.

Teaching objectives

For an introductory session

- To describe the frequency, trends, and burden of main respiratory diseases nationally and globally.
- To describe the major risk factors of chronic airway diseases, with particular attention to smoking, occupation, air pollution, and inhaled allergens, and the currently available prevention strategies.

For a standard course in a master's of public health or clinical research

- To describe the frequency, trends, and burden of main respiratory diseases.
- To describe the major risk factors of chronic airway diseases, with particular attention to smoking, occupation, air pollution, and inhaled allergens, and how their effects can be prevented.
- To understand the main issues of the respiratory epidemiology of asthma and COPD with respect to prevalence and incidence, disease definition, natural history, etiology, prevention, and control.
- To become familiar with the landmark studies of respiratory epidemiology.
- To develop a policy-oriented perspective based on an understanding of how epidemiology contributes to the development of strategies for the control and prevention of respiratory diseases, including issues like evidence-based medicine and the effectiveness of the available interventions.

Teaching content (including a syllabus)

Frequency and burden of disease

The first section in any programme of respiratory epidemiology is likely to include a short presentation of the frequency, distribution, and burden of respiratory diseases in the population. Emphasis should be given to those entities that are more relevant in terms of mortality and morbidity. COPD is both an important cause of morbidity and mortality and a disease for which a more sustained rise is predicted to occur. In the last Global Burden Disease Study, COPD was considered the ninth leading cause of disability-adjusted life years (DALYs) worldwide (seventh in Western Europe and second in high-income North America) in 2010, whereas asthma, which affects both children and adults, is the twenty-sixth cause of DALYs worldwide (Murray et al. 2012).

Asthma prevalence has been increasing worldwide during the last decades—an increase that has stimulated a large amount of epidemiological research,

although the increase has recently begun to plateau in the areas with the highest prevalence (Asher et al. 2006). The analysis of time trends in asthma can be used to introduce the issue of age, cohort, and period effects (Sunyer et al. 1999). Assessment of the disease distribution at the national and international levels is a relatively new area that has been covered by several large international studies such as the International Study of Asthma and Allergies in Childhood (ISAAC) for asthma in children (Asher et al. 2006), the European Community Respiratory Health Survey (ECRHS) for asthma in adults (Janson et al. 2001), and the Burden of Obstructive Lung Disease study (BOLD) for COPD (Mannino and Buist 2007). In the case of asthma, there were substantial hopes that tracing its geographical distribution worldwide would contribute to the understanding of its etiology. To date, however, there is no single etiological theory that provides a reasonable account for the observed geographical patterns.

Teaching initiatives in respiratory epidemiology need to identify the epidemiological peculiarities of the field in terms of both concepts and methods and address them in a way that provides the student with the appropriate knowledge and skills to understand current research. One of these particularities is the intermittent course of asthma (Strachan et al. 1996). Because of its intermittent course, the concept of 'current asthma' is frequently used, and the definition needs to be well understood. Questionnaires usually inquire about asthma by asking about the presence of respiratory symptoms in the last twelve months. Incident asthma usually refers to people who have reported respiratory symptoms during the last year for the first time in their life. In longitudinal studies with sufficient repeated measurements, it is possible to estimate the incidence, remission, and persistence (Sears et al. 2003), although misclassification of disease status in repeated observations may bias incidence, a problem that could be circumvented by using the net change in prevalence as an alternative to incidence (Chinn et al. 2004).

Disease definition

GINA defines asthma as follows:

> Asthma is a heterogeneous disease, usually characterized by chronic airway inflammation. It is defined by the history of respiratory symptoms such as wheeze, shortness of breath, chest tightness and cough that vary over time and in intensity, together with variable expiratory airflow limitation (GINA 2014).

Unfortunately, a pragmatic measure of such definition is difficult to achieve, and asthma is usually defined in epidemiological studies by means of respiratory symptoms reported in questionnaires. There are several questionnaires that are widely used, in particular, the ones used for asthma by the ECRHS for

adults (Janson et al. 2001) and the ISAAC for children (Asher et al. 1995). Questionnaire-based definitions of asthma have been traditionally validated using bronchial hyperresponsiveness (BHR) as a reference criterion (Burney et al. 1989). Video questionnaires were developed for children for use in the ISAAC (Crane et al. 1995). As in most diseases, however, any definition of asthma will involve a degree of misclassification, and the most appropriate choice will depend on the aim of the study. Definitions that maximize the Youden index (sensitivity + specificity − 1) are more appropriate for prevalence studies, and those that maximize the positive predictive value are more appropriate for etiological studies (Pekkanen and Pearce 1999). The latter type of definition can be achieved by combining the presence of symptoms and BHR into a single definition of asthma. Bronchial responsiveness is the physiological tendency of the airways to react to several physical (like exercise and cool air) or chemical stimuli and can be reversed by the administration of bronchodilators. BHR is an exaggerated bronchoconstrictive response to external stimuli, and its presence can be considered either a susceptibility factor or a characteristic of asthma. Assessment of BHR is complex and usually involves the inhalatory administration of drugs like metacholine or histamine. The other important susceptibility factor of asthma is atopy. Atopy is the tendency to develop specific IgE antibodies to allergens and can be assessed by measuring the level of circulating specific IgE antibodies in the blood or by the skin test method. Atopy is closely associated with asthma, although the causal direction of the association is not well established (Pearce et al. 1999). Asthma is frequently classified as atopic or non-atopic asthma.

Defining COPD involves different issues. The most common definition of COPD is the one proposed by the GOLD initiative, which is based on the so-called fixed ratio of lung function parameters. Ratios (FEV_1/FVC) are determined with stages of severity defined according to the level of observed versus predicted FEV_1, and both parameters are obtained after a bronchodilator test. The administration of a bronchodilator is used to exclude the intrinsic short-term variability in lung function. The main difficulty in defining COPD is that the prevalence of the disease is highly dependent on the definition used, which can be arbitrary (Celli et al. 2003). Because lung function level is age dependent and the fixed ratio can underdiagnose COPD in younger people and overdiagnose in the elderly, the alternative use of the lower limit of normality has been recommended (Hnizdo et al. 2006). Although it is beyond the scope of this chapter, the use of lung function measures, including the contentious use of reference equations for spirometry, is a central aspect of respiratory epidemiology (Pistelli et al. 2007). The interpretation of pulmonary functions tests usually involves a comparison of the observed versus predicted values (Pellegrino et al.

2005). The predicted values are usually obtained from an external 'healthy' population with similar anthropometrics and ethnic backgrounds, whereas most reference equations are obtained from cross-sectional data and thus are subject to a cohort effect.

During the last decade, the emergence of genetic studies that have frequently led to inconsistent results has increased the interest in understanding the heterogeneity of the asthma and COPD phenotypes. The hope is that modelling this heterogeneity could help identify its genetic and environmental determinants; such a hope is based on the assumption that the heterogeneity is the result of the different genetic and environmental determinants. In the case of COPD, the view that current definitions based on lung function may be too narrow is gaining attention, and there is an increasing interest in measuring emphysema with the new helicoidally computerized tomography. A composite index including BMI, dyspnoea, and exercise capacity, in addition to lung function, has proven a better predictor of mortality than lung function alone (Celli et al. 2004). However, changing a disease definition is not trivial and certainly involves substantial complexity. On the one hand, there is a philosophical level of complexity, since individual diseases are not observable entities by themselves but complex constructs in which the definition is dependent on the available knowledge and purpose. On the other hand, there is an empirical level of complexity since, among the many different possible definitions, each one will have advantages and disadvantages. The choice here depends on the purpose of the definition. In the case of asthma, the possibility of developing a definition based on a continuous measure instead of a dichotomic one, as advocated by Geoffrey Rose (1995), has been recently approached by additively scoring a number of asthma-related respiratory symptoms (Sunyer et al. 2007). In the case of COPD, one can see FEV_1 or the FEV_1/FVC ratio as a continuous measure of COPD, to avoid the difficulties involved in any threshold-based categorical definition. In this respect, a teaching programme in respiratory epidemiology provides a good opportunity to familiarize students with the complexity of defining diseases and the different alternatives.

Etiology

Etiology is a central topic of interest in epidemiology. Any course of respiratory epidemiology will need to pay substantial attention to etiological research and to present related concepts like causal models, multicausality, and attributable fractions. Respiratory epidemiology provides one of the best paradigms of causal risk factors in non-infectious chronic diseases—the role of smoking in the development of COPD and cancer of the lung—a paradigm which fits narrowly the Bradford–Hill criteria of causation, as shown in a landmark study by

Doll and Hill (1964). Because of the strong evidence that smoking is a major cause of both COPD and lung cancer, respiratory physicians have become some of the lead participants in the fight against smoking.

In developing countries, indoor chronic exposure to biomass fuel is considered a cause of COPD and is likely to explain a substantial proportion of COPD in women and non-smokers. In addition to smoking and biomass, other risk factors have been associated with the development of COPD, including traffic and other outdoor pollution, second-hand smoke, and dietary factors (Eisner 2010). Occupational exposures (such as mining, and dust exposures in textile industries) have also been linked to the development of COPD, with a recent review concluding that occupational exposures could account for up to 15 per cent of COPD (Balmes et al. 2003). Other factors for which there is some causal evidence are socio-economic status (Prescott and Vestbo 1999), a diet poor in antioxidants (Romieu and Trenga 2001; Varraso et al. 2007), and a lack of physical activity (García-Aymerich et al. 2007). Both in developed and developing countries, an almost ubiquitous risk factor in urban areas is traffic-related, outdoor air pollution. Several longitudinal studies have suggested this as the cause of the increase in the risk of lung function decline and thus likely to contribute to the development of COPD (Gauderman et al. 2004). Because COPD is usually diagnosed in people over 60 and because smoking usually begins after adolescence, there is a tendency to think that risk factors of COPD play their role during adulthood. There is evidence, however, that exposures occurring early in life (even in utero) may affect lung growth and subsequently result in reduced lung function development (Shaheen et al. 1998). In a longitudinal study, subjects with early life respiratory risk factors, including maternal asthma, paternal asthma, childhood asthma, maternal smoking, and childhood respiratory infections, had an increased COPD risk with an impact that was as large as that of heavy smoking (Svanes et al. 2010). This study can be used to introduce or reinforce the students' knowledge regarding Barker's hypothesis that chronic diseases in adults may at least partially result from the biological programming effects of prenatal environmental exposures (Baker 2000), and the importance of life course epidemiology.

The etiology of asthma is by far less understood and more difficult to investigate than COPD (Eder et al. 2006). During the 1980s, the allergic theory of asthma, that is, that exposure to allergens was a major cause of asthma development, was the more prevalent view regarding the etiology of asthma. During the last decade, several studies have failed to replicate early findings that exposure to house-dust mite allergens in early life is associated with an increased risk of developing asthma (Lau et al. 2000). There is growing evidence that the lungs' response to an inhaled allergen is more complex than was initially

believed and that the lungs can have a complex dose-response function—including a tolerance of high doses of an allergen, and differences in the behavioural reaction to different allergens. The latter may explain why consistent evidence of a causal relationship between exposure to inhaled allergens and the development of asthma has not been established (Antó 2004; Torrent et al. 2007). Several randomized controlled trials of primary prevention during pregnancy and early life have included interventions to reduce exposure to house-dust mite allergens. Since most of these studies had inconclusive results and presented issues such as dietary restriction in the mothers' diets or short follow-up, their interpretations remain difficult (Simpson and Custovic 2009).

The role of infection in the development of asthma is not yet understood. Infections caused by the respiratory syncytial virus have been shown to increase the risk of childhood asthma (Stein et al. 1999), whereas other infections such as measles may decrease the risk (Shaheen et al. 1996). A similar pattern has been described for parasitic infections (Leonardi-Bee et al. 2006). The 'hygiene hypothesis' was formulated by Strachan (2000) to propose that early exposure to infection or to microbial-related products protects against developing allergy and probably asthma. Reductions in the exposure to microbial products at the population level could partially account for the generalized increase in the prevalence of asthma. Evidence in support of the hygiene hypothesis has come from cross-sectional studies conducted in farming families, where infants exposed to stables during their gestational period and first months of life were found to have an exceedingly low risk of allergy and atopic wheezing (Braun-Fahrländer et al. 2002). Other exposures that have been related to a protective effect of allergy and asthma in the context of the hygiene hypothesis are older siblings, day care, pet dogs, and infection by the hepatitis A virus (Strachan 2000). Strong evidence supporting the hygiene hypothesis, however, is yet to come, and available evidence is by far more consistent for allergic rhinitis and sensitization than for asthma. An evolved version of the hygiene hypothesis is the theory that reduction in environmental biodiversity has led to depletion of protective immunomodulatory microbes in human organs like the skin and the gut and contributed to the epidemic increase of allergy (Haahtela et al. 2013).

Perhaps one of the best models of the etiology of asthma has been facilitated by studies of occupational asthma. In this field there is wide evidence from longitudinal studies that exposure to allergens involved in baking and in handling animals in research laboratories is associated with an increased risk of developing new-onset asthma (Cullinan et al. 1999, 2001). So far, approximately 300 different agents have been reported as capable of causing asthma in occupational environments. Studies conducted in general populations have indicated

that 12–25 per cent of incident cases of asthma in young adults may be attributable to occupation (Kogevinas et al. 2007).

Beyond the etiological theories of asthma reviewed above, there is a wide range of environmental exposures that have been linked to the development of asthma in epidemiological studies. In the last update of the GINA report, moderate evidence of causality is believed to exist for the following environmental factors: indoor and outdoor allergens (domestic mites, furred animals, cockroach allergens, fungi, moulds, yeasts, pollens), infections (predominantly viral), occupational sensitizers, tobacco smoke, indoor and outdoor air pollution, and diet (Global Initiative for Asthma 2014). Among non-environmental risk factors, there is substantial consistence across studies that gender, obesity, and a family history of asthma are associated with an increased risk of asthma. The association between obesity and asthma has been reproduced in both cross-sectional and longitudinal studies (Chinn 2006).

The role of genomic factors

Genetic determinants of chronic obstructive respiratory diseases are increasingly being investigated as relevant determinants of the development of asthma and COPD and related traits such as lung function and atopy. In recent years, a better understanding of the human genome, together with the availability of high-throughput technologies, has led the genome-wide association approach to become the standard. By 2000, a total of twenty-two genome-wide association approach studies covering respiratory outcomes such as pulmonary function measures, onset of asthma, and susceptibility to COPD had been published (Todd et al. 2011). Unfortunately, a coherent picture is still missing, and unexplained heritability is still the predominant finding. In these studies, rare variants give the highest risk estimates but are limited to small subgroups; thus, these studies indicate a complex origin of asthma that, if truly causal, may involve hundreds of variants that are either population, family, or individual specific (Wjst et al. 2013). Two meta-analyses identified four chromosomal regions consistently associated with development of asthma (Moffat et al. 2007; Torgerson et al. 2011). One of these meta-analyses found that multiple markers on chromosome 17q21 were strongly and reproducibly associated with childhood onset asthma and, further, that genetic variants regulating *ORMDL3* expression may be determinants of susceptibility to childhood asthma (Moffat et al. 2007). In addition, genes that are associated with asthma subphenotypes such as lung function, biomarker levels, and asthma therapeutic responses can provide insight into mechanisms of asthma progression and severity. Currently, new genetic studies using complete sequencing techniques and more detailed phenotyping are being developed and are likely to substantially improve our

currently limited understanding of the genetic regulation of respiratory diseases. In addition to the structural genetic variants, epigenetic mechanisms are likely to play a role in asthma and COPD since epigenetic mechanisms such as DNA methylation, histone modification, and microRNA may modulate environmental factors and influence disease development (Kabesch and Adcock 2012). With the rapid increase in the availability and cost of different 'omic' technologies, complex integrative studies, based on the model developed by systems biology, are gaining interest in respiratory medicine research.

Of special interest to epidemiologists is the hope that a better understanding of genetic determinants may facilitate the assessment of gene–environment interactions and lead to better etiological evidence. Among the available examples, one that can be used for teaching purposes is the interaction between gene variants (i.e. *GSTM1* null and *GSTP1* valine/valine) and exposure to outdoor ozone, as it is linked to an increase in the risk of respiratory symptoms in asthmatic children in Mexico City (Romieu et al. 2006). There is also the complex interaction between *CD14* variants and exposure to endotoxins, as different alleles seem to have opposite effects depending on the levels of endotoxin (Martínez 2007). Despite these promising results, many positive findings have not been replicated, and overall the results from the literature remain inconclusive (London and Romieu 2009).

Natural history

Asthma and COPD are chronic diseases that evolve with intermittent periods of disease worsening (i.e. exacerbations). There are several reasons for including the natural history of asthma and COPD in a course of respiratory epidemiology. One reason is that the risk factors that are relevant for the incidence of the disease may not be the same as those that influence the progression of the disease. Authors have referred to these factors using different terms, such as contributing or exacerbating factors. In addition, asthma can present remissions. Understanding what the determinants of remissions are may eventually open new ways of treatment and prevention. Finally, both COPD exacerbations and asthma attacks are acute events that require specific epidemiological designs to be investigated, such as time-series and case cross-over designs (Sunyer et al. 2000).

Among the risk factors of COPD exacerbation, the ones for which there is more solid evidence are infection, traffic-related air pollution, and lack of physical activity. Regarding asthma, the best evidence is that exposure to inhaled allergens in previously sensitized asthmatics may trigger asthma symptoms and exacerbations. Other environmental factors that increase asthma morbidity are active and passive exposure to tobacco smoke and traffic-related air pollution.

In some circumstances, asthma exacerbations can be experienced simultaneously by a large number of asthmatics in a given community. Repeated epidemics of asthma affecting more than 1,000 people and producing approximately twelve deaths were registered and studied in the city of Barcelona. By combining case-control and time-series studies, it was shown that these epidemics were due to the inhalation of soybean dust spread over the city under appropriate meteorological conditions during the unloading of soybean from ships into a harbour silo that lacked appropriate air filters (Antó et al. 1989). A subsequent intervention study showed that asthma epidemics disappeared after the installation of appropriate filters into the silo (Antó et al. 1993). Other asthma outbreaks have also occurred in close association with thunderstorms (Antó and Sunyer 1997).

Other issues that are relevant aspects of the natural history of chronic respiratory diseases are prognosis and severity. These are complex issues that cannot be easily covered in a general course but that may need some attention depending on the aims of the course and the interests of the participants. A study of the predictive indexes for COPD (Puhan et al. 2009) and of the methods to assess asthma control (Bateman et al. 2004) can be used to illustrate the application of standard epidemiological methods to respiratory diseases.

Treatment

Epidemiology is aimed at studying the determinants of health and disease, including the health services at the community level and has a leading role in teaching the principles of medical treatments, including the concepts and methods for assessing efficacy, effectiveness, and efficiency. In the case of respiratory disease, there are good opportunities to teach students the use of epidemiology for this purpose. In several countries, waves of asthma mortality during the 1960s, 1970s, and 1980s were identified. In New Zealand a case-control study of asthma deaths showed that asthma mortality was associated with the use of fenoterol, a metered dose inhaler beta-agonist (Crane et al. 1989). After publication of the case-control study, the New Zealand asthma death rate fell by half (Pearce et al. 1995). The worldwide increase in the prevalence of asthma during the 1980s has favoured increased attention to the distribution of appropriate treatment and control at the community level. In a similar way, data from the ECRHS were used to assess asthma control according to the GINA guidelines. In this study, six out of seven European asthmatic adults using inhaled steroids in the last year did not achieve good disease control, and the large majority of subjects with poorly controlled asthma were using anti-asthma drugs in a suboptimal way (Cazzoletti et al. 2007). Although these types of cross-sectional studies do not allow for the establishment of a causal

relationship between the level of treatment and the level of control, they are useful to estimate the distribution of treatment and the control of asthma at the community level and to inform health-care policy efforts.

Epidemiology has an important role in assessing the beneficial and harmful effects of medical care, and in this respect there are some issues, including the role of the randomly controlled trial, that can be included in teaching programmes of respiratory epidemiology. In recent years there has been a debate about the role of inhaled steroids in the reduction of mortality in COPD patients. Although some observational studies did show a reduction in mortality among COPD patients receiving inhaled steroids, other studies controlling for the presence of immortal time bias found no reduction in mortality (Suissa 2008).

Prevention and control

Primary prevention of respiratory diseases is based on the reduction and, if possible, the elimination of the exposure to established risk factors. Special attention should be given to the control and elimination of smoking. As mentioned earlier, the fight against smoking is a fundamental concept of respiratory medicine and one of its strongest links to public health. Most national and international societies of respiratory diseases play an active role against smoking and have become important allies of public health authorities. Prevention of smoking is considered the main strategy for primary prevention of COPD ever since early studies in the 1970s established a close causal link. Recent evidence has reinforced this view, showing that maternal smoking (in particular during pregnancy) is associated with lower lung function in childhood (Cook et al. 1998). There is abundant evidence about individual smoking cessation programmes that, despite a relatively low rate of sustained quitting, are considered to be cost-effective in the context of the National Health Service in the UK (US Centers for Disease Control and Prevention 1995). Several programmes of banning smoking in public places have recently been shown to be useful (Schmidt 2007). Reduction of air pollution is also an important strategy for primary prevention of respiratory diseases. Impact assessment studies have reported that a substantial amount of mortality and morbidity due to respiratory diseases is attributable to exposure to outdoor air pollution (Kunzli et al. 2000). In Dublin, after a ban on the use of coal, a reduction of mortality due to cardiovascular and respiratory diseases was observed (Clancy et al. 2002).

In the case of asthma, primary prevention efforts have focused on high-risk children (high risk usually being assessed through family antecedents of asthma, allergy, or sensitization). In these children, primary prevention has the purpose of preventing the development of asthma and other allergic diseases

such as rhinitis or dermatitis. There are several randomly controlled trials reported in the literature that have included food allergen avoidance, and intervention to reduce the levels of house-dust mites and other allergens using combined strategies. Even though some trails have reported a substantial reduction of asthma and allergic diseases in the intervention group, its interpretation requires caution, and recommendations cannot be formulated from a public health perspective (Simpson and Custovic 2009).

Another potential strategy for primary prevention of respiratory diseases is the reduction of occupational exposure to airborne substances. In the case of occupational asthma, reduction of exposure to specific asthmagens offers opportunities for primary prevention. Some exposures that have been partially controlled for by changes in the manufacturing or industrial processes are latex and detergents (Cullinan et al. 2000; LaMontagne et al. 2006). There is little evidence, however, of successful primary prevention of occupational asthma, and the selection of new employees that are not at high risk of asthma is inefficient and unethical. Because most cases of occupational asthma came from those at low risk, who are the largest fraction in the general population, approaching prevention through strategies targeted to workers who are allergic or at higher risk of asthma is not useful (Cullinan et al. 2003; Nicholson et al. 2005). Secondary prevention at the level of the population usually includes an early diagnosis through a screening phase, plus an intervention when the diagnosis is confirmed. This type of prevention has not achieved an established role in chronic obstructive airway diseases. In the case of occupational asthma, early diagnosis followed by the removal of exposure in high-risk occupations has been shown to increase the remission of asthma or to improve its evolution (Cullinan et al. 2000). Regarding the secondary prevention of allergic diseases, there has been a lot of debate about the role of specific allergen avoidance measures. Despite the evidence that, when vigorously applied, such measures may reduce the risk of sensitization and symptoms of allergic respiratory disease, their effectiveness and cost-effectiveness has not been satisfactory demonstrated (Bush 2008). The available evidence suggests that interventions in children may be associated with some beneficial effect on asthma control but no conclusive evidence exists regarding rhinitis or eczema. Conversely, there is little evidence to support the recommendation of allergen avoidance methods in adults with asthma and rhinitis (Marinho et al. 2006). Regarding COPD, the possibility of early diagnosis through repeated spirometry in smokers has recently gained attention, the rationale being that an early spirometric diagnosis of COPD could be followed by more effective response to anti-smoking advice. Unfortunately, there is no evidence available to know whether screening for COPD using spirometry reduce morbidity and mortality, and so far spirometry

does not seem to increase smoking cessation rates. For these reasons, the US Preventive Task Force recommended against screening adults for COPD using spirometry (US Preventive Services Task Force 2008). Finally, tertiary prevention consists of improving the outcomes of the disease through preventive interventions once the disease is already established. Tertiary prevention can be best understood in COPD, for which smoking cessation, influenza vaccination, and respiratory rehabilitation programmes have been shown to be cost-effective interventions (Global Initiative for Chronic Obstructive Lung Disease 2014).

Teaching methods and format

Combining different teaching methodologies is a useful strategy for teaching respiratory epidemiology. Lectures should be allocated to address general issues and to introduce the more relevant conceptual and methodological aspects. When used appropriately, lectures are useful to attract the attention of students to the general concepts and issues, as well as the landmark studies. Including information from recent issues that have gained wide visibility in mass media is also useful. An important component of teaching programmes should be based on problem-solving strategies. A simple tool is to use case studies based on critical reading of relevant research articles, with the aid of a list of pre-established questions focusing the students' attention to the more relevant issues. Depending on the programme and the time available, case studies can be managed with two to four hours of work in parallel sessions with small groups (four to six) of students, followed by a session together. Another learning method is conducting a short piece of research, which, depending on the programme characteristics, can range from developing a short research protocol to a complete research project. It is also important to include appropriate teaching materials such as selected references comprising original research articles, reviews, guidelines, and other scientific and technical documents. Books may also play a role, whether they are general public health textbooks that may contain a chapter on respiratory disease (Detels et al. 2004) or specialized books (Stockley et al. 2007), with some being nicely targeted to teaching purposes (Pearce et al. 1998). The European Respiratory Society (ERS) provides an online respiratory epidemiology course which has been granted six continuing medical education credits by the European Board for Accreditation in Pneumology.

Evaluation: assessing students' achievements and teaching performance

How to assess each student's achievements is a critical step in any programme. The assessment should be commensurate to the programme characteristic and

should be comprehensive in terms of its different components. Appropriate combinations of written exams, based on either multiple-choice or short questions, plus an evaluation of participation in the case studies and the research work is likely to provide an overall appropriate view of the student's achievements. Participation in the working groups is difficult to evaluate but considering attendance and using self-rated questionnaires can be useful.

References

Antó, J. M. (2004) The causes of asthma: the need to look at the data with different eyes. *Allergy*, **59**: 121–3.

Antó, J. M. et al. (1993) Preventing asthma epidemics due to soybeans by dust-control measures. *New England Journal of Medicine*, **329**: 1760–3.

Antó, J. M. and Sunyer, J. (1997) Thunderstorms: a risk factor for asthma attacks. *Thorax*, **52**: 669–70.

Antó, J. M., Sunyer, J., Rodriguez-Roisin, R., Suarez-Cervera, M., and Vazquez, L. (1989) Community outbreaks of asthma associated with inhalation of soybean dust. Toxicoepidemiological Committee. *New England Journal of Medicine*, **320**: 1097–102.

Asher, M. I. et al. (1995) International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *European Respiratory Journal*, **3**: 483–91.

Asher, M. I., Montefort, S., Björkstén, B., Lai, C. K., Strachan, D. P., Weiland, S. K., and Williams, H. (2006) ISAAC Phase Three Study Group. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet*, **368**: 733–43.

Barker, D. J. P. (2000) *Fetal Origins of Cardiovascular and Lung Disease*, vol 151. New York: Marcel Dekker Inc.

Balmes, J., Becklake, M., Blanc, P., Henneberger, P., Kreiss, K., Mapp, C., Milton, D., Schwartz, D., Toren, K., and Viegi, G. (2003) Environmental and Occupational Health Assembly, American Thoracic Society. American Thoracic Society Statement: Occupational contribution to the burden of airway disease. *American Journal of Respiratory Critical Care Medicine*, **167**: 787–97.

Bateman, E. D., Boushey, H. A., Bousquet, J., Busse, W. W., Clark, T. J., Pauwels, R. A., and Pedersen, S. E; GOAL Investigators Group. (2004) GOAL Investigators Group. Can guideline-defined asthma control be achieved? The Gaining Optimal Asthma Control study. *American Journal of Respiratory Critical Care Medicine*, **170**: 836–44.

Braun-Fahrländer, C. et al. (2002) Environmental exposure to endotoxin and its relation to asthma in school-age children. *New England Journal of Medicine*, **347**: 869–77.

Burney, P. G., Chinn, S., Britton, J. R., Tattersfield, A. E., and Papacosta, A. O. (1989) What symptoms predict the bronchial response to histamine? Evaluation in a community survey of the bronchial symptoms questionnaire (1984) of the International Union Against Tuberculosis and Lung Disease. *International Journal of Epidemiology*, **18**: 165–73.

Bush, R. K. (2008) Indoor allergens, environmental avoidance, and allergic respiratory disease. *Allergy and Asthma Proceedings*, **29**: 575–9.

Cazzoletti, L. et al. (2007) Therapy and Health Economics Group of the European Community Respiratory Health Survey. Asthma control in Europe: a real-world evaluation based on an international population-based study. *Journal of Allergy and Clinical Immunology*, **120**: 1360–7.

Celli, B. R., Cote, C. G., Marin, J. M., Casanova, C., Montes de Oca, M., Mendez, R. A., Pinto Plata, V., and Cabral, H. J. (2004) The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *New England Journal of Medicine*, **350**: 1005–12.

Celli, B. R., Halbert, R. J., Isonaka, S., and Schau, B.. (2003) Population impact of different definitions of airway obstruction. *European Respiratory Journal*, **22**: 268–73.

Chinn, S. (2006) Obesity and asthma. *Paediatric Respiratory Reviews*, **7**: 223–8.

Chinn, S. et al. (2004) Increase in diagnosed asthma but not in symptoms in the European Community Respiratory Health Survey. *Thorax*, **59**: 646–51.

Clancy, L., Goodman, P., Sinclair, H., and Dockery, D. W. (2002) Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet*, **360**: 1210–4.

Cook, D. G., Strachan, D. P., and Carey, L. M. (1998) Health effects of passive smoking. 9: Parental smoking and spirometric indices in children. *Thorax*, **53**: 884–93.

Crane, J., Beasley, R., Stewart, B., Shaw, R., Pearce, N., and Burgess, C. (1995) The international asthma video questionnaire for measuring asthma prevalence in different populations. *International Archives of Allergy and Immunology*, **107**: 450–1.

Crane, J., Flatt, A., Jackson, R., Ball, M., Pearce, N., Burgess, C., Kwong, T., and Beasley, R. (1989) Prescribed fenoterol and death from asthma in New Zealand, 1981–83: case-control study. *Lancet*, **333**: 917–22.

Cullinan, P., Cook, A., Gordon, S., Nieuwenhuijsen, M. J., Tee, R. D., Venables, K. M., McDonald, J. C., and Taylor, A. J. (1999) Allergen exposure, atopy and smoking as determinants of allergy to rats in a cohort of laboratory employees. *European Respiratory Journal*, **13**: 1139–43.

Cullinan, P., Cook, A., Nieuwenhuijsen, M. J., Sandiford, C., Tee, R. D., Venables, K. M., McDonald, J. C., and Newman Taylor, A. J. (2001) Allergen and dust exposure as determinants of work-related symptoms and sensitization in a cohort of flour-exposed workers; a case-control analysis. *Annals of Occupational Hygiene*, **45**: 97–103.

Cullinan, P., Harris, J. M., Newman Taylor, A. J., Hole, A. M., Jones, M., Barnes, F., and Jolliffe G. (2000) An outbreak of asthma in a modern detergent factory. *Lancet*, **356**: 1899–900.

Cullinan, P., Tarlo, S., and Nemery, B. (2003) The prevention of occupational asthma. *European Respiratory Journal*, **22**: 53–60.

Detels, R., McEwen, J., Beaglehole, R., and Tanaka, H. eds. (2004) *Oxford Textbook of Public Health* (4th edn). New York: Oxford University Press.

Doll, R. and Hill, A. B. (1964) Mortality in relation to smoking: ten years' observations of British doctors. *British Medical Journal*, **1**: 1399–410.

Eder, W., Ege, M. J., and von Mutius, E. (2006) The asthma epidemic. *New England Journal of Medicine*, **355**: 2226–35.

Eisner, M. D., Anthonisen, N., Coultas, D., Kuenzli, N., Perez-Padilla, R., Postma, D., Romieu, I., Silverman, E. K., Balmes, J. R., and the Committee on Nonsmoking COPD, Environmental and Occupational Health Assembly. (2010) An official

American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, **182**: 693–718.

García-Aymerich, J., Lange, P., Benet, M., Schnohr, P., and Antó, J. M. (2007) Regular physical activity modifies smoking-related lung function decline and reduces risk of chronic obstructive pulmonary disease: a population-based cohort study. *American Journal of Respiratory Critical Care Medicine*, **175**: 458–63.

Gauderman, W. J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., Margolis, H., Bates, D., and Peters, J. (2004) The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine*, **351**: 1057–67.

Global Initiative for Asthma. (2014) *Global Initiative for Asthma (GINA) 2014*. <http://www.ginasthma.com>, accessed 12 December 2014.

Global Initiative for Chronic Obstructive Lung Disease. (2014) *Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2014*. <http://www.goldcopd.org>, accessed 12 December 2014.

Haahtela, T., Holgate, S., Pawankar, R., Akdis, C. A., Benjaponpitak, S., Caraballo, L., Demain, J., Portnoy, J., and von Hertzen, L. (2013) WAO Special Committee on Climate Change and Biodiversity. The biodiversity hypothesis and allergic disease: world allergy organization position statement. *World Allergy Organ Journal*, 6: 3.

Hnizdo, E., Glindmeyer, H. W., Petsonk, E. L., Enright, P., and Buist, A. S. (2006) Case definitions for chronic obstructive pulmonary disease. *COPD*, **3**: 95–100.

Janson, C. et al. (2001) European Community Respiratory Health Survey II. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. *European Respiratory Journal*, **18**: 598–611.

Kabesch, M. and Adcock, I. M. (2012) Epigenetics in asthma and COPD. *Biochimie*, **94**: 2231–41.

Kogevinas, M. et al. (2007) Exposure to substances in the workplace and new-onset asthma: an international prospective population-based study (ECRHS-II). *Lancet*, **370**: 336–41.

Künzli, N. et al. (2000) Public health impact of outdoor and traffic-related air pollution: a tri-national European assessment. *Lancet*, **356**: 795–801.

LaMontagne, A. D., Radi, S., Elder, D. S., Abramson, M. J., and Sim, M. (2006) Primary prevention of latex related sensitisation and occupational asthma: a systematic review. *Occupational and Environmental Medicine*, **63**: 359–64.

Lau, S., Illi, S., Sommerfeld, C., Niggemann, B., Bergmann, R., von Mutius, E., Wahn, U., and the Multicentre Allergy Study Group. (2000) Early exposure to house-dust mite and cat allergens and development of childhood asthma: a cohort study. *Lancet*, **356**: 1392–7.

Leonardi-Bee, J., Pritchard, D., and Britton, J. (2006) Asthma and current intestinal parasite infection: systematic review and meta-analysis. *American Journal of Respiratory Critical Care Medicine*, **174**: 514–23.

London, S. J. and Romieu, I. (2009) Gene by environment interaction in asthma. *Annual Review of Public Health*, **30**: 55–80.

Mannino, D. M. and Buist, A. S. (2007) Global burden of COPD: risk factors, prevalence, and future trends. *Lancet*, **370**: 765–73.

Marinho, S., Simpson, A., and Custovic, A. (2006) Allergen avoidance in the secondary and tertiary prevention of allergic diseases: does it work? *Primary Care Respiratory Journal*, **15**: 152–8.

Martínez, F. D. (2007) CD14, endotoxin, and asthma risk: actions and interactions. *Proceedings of the American Thoracic Society*, **4**: 221–5.

Moffatt, M. F. et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, **448**: 470–3.

Murray, J. L. et al. (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**: 2197–223.

Nicholson, P. J., Cullinan, P., Taylor, A. J., Burge, P. S., and Boyle, C. (2005) Evidence based guidelines for the prevention, identification, and management of occupational asthma. *Occupational and Environmental Medicine*, **62**: 290–9.

Pearce, N., Beasley, R., Burgess, C., Crane, J. (1998) *Asthma Epidemiology: Principles and Methods*. New York: Oxford University Press.

Pearce, N., Beasley, R., Burgess, C., and Jackson, R. (1995) End of the New Zealand asthma mortality epidemic. *Lancet*, **345**: 41–4.

Pearce, N., Pekkanen, J., and Beasley, R. (1999) How much asthma is really attributable to atopy? *Thorax*, **3**: 268–72.

Pekkanen, J. and Pearce, N. (1999) Defining asthma in epidemiological studies. *European Respiratory Journal*, **14**: 951–7.

Pellegrino, R. et al. (2005) Interpretative strategies for lung function tests. *European Respiratory Journal*, **26**: 948–68.

Pistelli, F., Bottai, M., Carrozzi, L., Baldacci, S., Simoni, M., Di Pede, F., and Viegi, G. (2007) Reference equations for spirometry from a general population sample in central Italy. *Respiratory Medicine*, **101**: 814–25.

Prescott, E. and Vestbo, J. (1999) Socioeconomic status and chronic obstructive pulmonary disease. *Thorax*, **54**: 737–41.

Puhan, M. A. et al. (2009) Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet*, **374**: 704–11.

Romieu, I., Ramirez-Aguilar, M., Sienra-Monge, J. J., Moreno-Macías, H., del Rio-Navarro, B. E., David, G., Marzec, J., Hernández-Avila, M., and London, S. (2006) GSTMI and GSTP1 and respiratory health in asthmatic children exposed to ozone. *European Respiratory Journal*, **28**: 953–9.

Romieu, I. and Trenga, C. (2001) Diet and obstructive lung diseases. *Epidemiologic Reviews*, **23**: 268–87.

Rose, G. (1995) *The Strategy of Preventive Medicine* (4th edn). Oxford: Oxford University Press.

Schmidt, C. (2007) A change in the air: smoking bans gain momentum worldwide. *Environmental Health Perspectives*, **115**: 412–15.

Sears, M. R., Greene, J. M., Willan, A. R., Wiecek, E. M., Taylor, D. R., Flannery, E. M., Cowan, J. O., Herbison, G. P., Silva, P. A., and Poulton, R. (2003) A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. *New England Journal of Medicine*, **349**: 1414–22.

Shaheen, S. O., Aaby, P., Hall, A. J., Barker, D. J., Heyes, C. B., Shiell, A. W., and Goudiaby, A. (1996) Measles and atopy in Guinea-Bissau. *Lancet*, **347**: 1792–6.

Shaheen, S. O., Sterne, A. J., Tucker, J. S., and Florey, C. D. (1998) Birth weight, childhood lower respiratory tract infection, and adult lung function. *Thorax*, **53**: 549–53.

Simpson, A. and Custovic, A. (2009) Prevention of allergic sensitization by environmental control. *Current Allergy and Asthma Reports*, **9**: 363–9.

Speizer, F. E. and Tager, I. N. (2007) 'Chronic respiratory disease epidemiology', in **W. W. Holland**, **J. Olsen**, and **C. du V. Florey**, eds, *The Development of Modern Epidemiology: Personal Reports from Those Who Were There*. New York: Oxford University Press, pp. 93–101.

Stein, R. T., Sherrill, D., Morgan, W. J., Holberg, C. J., Halonen, M., Taussig, L. M., Wright, A. L., and Martinez, F. D. (1999) Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years. *Lancet*, **354**: 541–5.

Stockley, R., Rennard, S., Rabe, K., and Celli, B. (2007) *Chronic Obstructive Pulmonary Disease*. Malden, MA: Blackwell Publishing.

Strachan, D. P. (2000) Family size, infection and atopy: the first decade of the 'hygiene hypothesis'. *Thorax*, **55**: S2–10.

Strachan, D. P., Buland, B. K., and Anderson, H. R. (1996) Incidence and prognosis of asthma and wheezing illness from early childhood to age 33 in a national British cohort. *British Medical Journal*, **312**: 1195–9.

Suissa, S. (2008) Immortal time bias in pharmaco-epidemiology. *American Journal of Epidemiology*, **167**: 492–9.

Sunyer, J., Antó, J. M., Tobias, A., and Burney, P., and the European Community Respiratory Health Study. (1999) Generational increase of self-reported first attack of asthma in fifteen industrialized countries: European Community Respiratory Health Study (ECRHS). *European Respiratory Journal*, **14**: 885–91.

Sunyer, J., Pekkanen, J., Garcia-Esteban, R., Svanes, C., Künzli, N., Janson, C., de Marco, R., Antó, J. M., and Burney, P. (2007) Asthma score: predictive ability and risk factors. *Allergy*, **62**: 142–8.

Sunyer, J., Schwartz, J., Tobías, A., Macfarlane, D., Garcia, J., and Antó, J. M. (2000) Patients with chronic obstructive pulmonary disease are at increased risk of death associated with urban particle air pollution: a case-crossover analysis. *American Journal of Epidemiology*, **151**: 50–6.

Svanes, C. et al. (2010) Early life origins of chronic obstructive pulmonary disease. *Thorax*, **65**: 14–20.

Todd, J. L., Goldstein, D. B,.Ge, D., Christie, J., and Palmer, S. M. (2011) The state of genome-wide association studies in pulmonary disease: a new perspective. *American Journal of Respiratory and Critical Care Medicine*, **184**: 873–80.

Torgerson, D. G. et al. (2011) Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nature Genetics*, **43**: 887–92.

Torrent, M., Sunyer, J., Garcia, R., Harris, J., Iturriaga, M. V., Puig, C., Vall, O., Anto, J. M., Newman Taylor, A. J., and Cullinan, P. (2007) Early-life allergen exposure and atopy, asthma, and wheeze up to 6 years of age. *American Journal of Respiratory Critical Care Medicine*, **176**: 446–53.

US Centers for Disease Control and Prevention. (1995) *Criteria for a Recommended Standard: Occupational Exposure to Respirable Coal Mine Dust*. Cincinnati, OH: National Institute of Occupational Safety and Health.

US Preventive Services Task Force. (2008) Screening for chronic obstructive pulmonary disease using spirometry: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*, **148**: 529–34.

Varraso, R., Fung, T. T., Barr, R. G,.Hu, F. B., Willett, W., and Camargo, C. A. Jr. (2007) Prospective study of dietary patterns and chronic obstructive pulmonary disease among US women. *American Journal of Clinical Nutrition*, **86**: 488–95.

Wjst, M., Sargurupremraj, M., and Arnold, M.. (2013) Genome-wide association studies in asthma: what they really told us about pathogenesis. *Current Opinion in Allergy and Clinical Immunology*, **13**: 112–18.

Chapter 22

Epidemiology of injuries

Eleni Petridou, Patricia Gerakopoulou, and Constantine N. Antonopoulos

Introduction to the epidemiology of injuries

Because epidemiology is the basic science of public health, and injuries are a major public health problem, injury epidemiology should be well represented in the corresponding field, and ideally injury prevention and control should be a priority in the allocation of public health resources. This is rarely the case, however, and the course instructor should try to explain to the students why injury epidemiology has received so little attention and why so few epidemiologists have chosen injury as the topic of their preference.

Epidemiology has been the principal discipline in preparing the grounds for the prevention of infectious diseases, several forms of cancer, and most types of cardiovascular disease. It should be admitted, however, that it has contributed relatively little towards the understanding of the causation and prevention of injuries. The teacher should put forward the following three possible issues, not necessarily mutually exclusive, that could explain the limited success of injury epidemiology in the past.

1. Currently used epidemiologic methods are not as appropriate for the investigation of injury occurrence as for the study of other diseases and causes of death. Many of the component causes of cardiovascular, neoplastic, or infectious disease are identifiable biologic entities, whereas the factors contributing to injuries are frequently of behavioural origin, poorly defined, and inadequately operationalized. In addition, there are distinct differences in the natural history of acute/chronic diseases and that of accidents and their consequences. Moreover, cohort studies, the method of choice in traditional epidemiology, are compromised in injury research by the inability to specify in advance the proximal behavioural or environmental exposures that may have triggered an accident. The course instructor could stimulate interest by asking which is the most appropriate design to estimate the magnitude of the risk in the case of drivers engaged in a distracting activity, such as an intense mobile-phone conversation, while driving. Given that the exposure is of a

transient nature, the case-crossover epidemiologic approach is best suited for studying injuries. In fact, the best information on road traffic injury prevention derives from case-control and case-crossover studies. Similarly, there is now an array of cohort studies that have provided valuable information on the component causes related to injury occurrence. However, case-control studies have inherent limitations in injury research, because instantaneous death, post-traumatic shock, or denial of responsibility on the part of those involved in an accident tend to create serious selection and information bias. These arguments, although legitimate, do not challenge the applicability of the discipline of epidemiology to injury research.

2. Injuries have not been recognized as a major public health problem until relatively recently. The lack of awareness about the magnitude of the public health impact of injuries is intuitively a source of frustration for both injury researchers and health and safety workers. The course instructor could, for example, ask whether students in the class are concerned about dioxin levels in food and then compare that answer with their level of concern on the impact of driving at high speeds; usually there is relatively much less concern about the latter, despite the fact that reducing average speed by 1 kilometre/hour would reduce the number of crashes by 4 per cent whereas quadrupling the average concentration of dioxin in the food would probably have no discernible effect on public health. In addition, the course instructor should point out that the high proportion of injuries among the young has a disproportionately large impact on life expectancy, a fact not immediately recognized by lay people, or even some health professionals.
3. Injury prevention research appears fragmented. Prevention of injuries, particularly of those caused by road traffic accidents, is frequently mixed with strong economic interests and power politics. Safety regulations and aspects of road construction, for instance, frequently overwhelm epidemiologic arguments supported by valid data and plain logic. The course instructor could ask the students to name potentially competing interests in road construction and audit procedures. This introduction, although defensive, may be more effective in capturing the interest of the students than the traditional one, which relies on a description of the magnitude of the injury problem but fails to address the applicability of the epidemiologic methods to injury etiology and prevention (Petridou 1997).

Teaching objectives

A course on general injury epidemiology for health professionals is usually of rather short duration. This is mainly because injury epidemiology is much

more specific for time, place, culture, technology, and activity than other branches of epidemiology. Moreover, the epidemiologic process for accidents (events) and injuries (lesions) is generally deductive rather than inductive, because causation of accidents and consequent injuries can be easily conceptualized in the context of the laws of physics and chemistry. Thus, to expand on factual information about when, where, and by what mechanism accidents have been caused in a particular place or at a particular time serves no general course objective.

The overall goal of a short course on the epidemiology of accidents and injuries is to document the importance of this field of epidemiology for the prevention of accidents and consequent injuries (Petridou 1997). More specifically, the course should aim to

- familiarize students with research on environmental or personal risk factors for injuries;
- enable students to assess the relative effectiveness of alternative intervention strategies for injury prevention; and
- enable students to evaluate the quality of health provision services and the impact of therapeutic interventions and rehabilitative processes.

Prevention of accidents corresponds to primary prevention, whereas prevention of death or serious injuries following an accident corresponds to secondary prevention. The overall goals can be accomplished by setting specific objectives. Thus, at the end of the course, the student is expected to be able to

- identify the critical role of injury epidemiology in the prevention and control of accidents and its distinct components;
- use the terms 'accident' or 'injury' correctly in particular situations;
- provide an adequate description for the causation of accidents in terms of energy transfer;
- name the major classification systems for accidents and injuries;
- describe the sources of data for accidents and injuries, as well as the strengths and limitations of these sources;
- apportion fatal injuries to major etiologic categories;
- delineate the contribution of descriptive data on the circumstances surrounding the accident to the prevention of injuries;
- appreciate the potential of sentinel injury data collection and early warning systems for the identification of hazardous products and the evaluation of injury prevention programmes;

- be familiar with the potential of relatively new study designs such as case-crossover or cluster randomized trials and choose the epidemiologic study design that best addresses the working hypothesis;
- point out the limitations of traditional study designs when applied to injury epidemiology and recognize the need to take into account complex exposure patterns in the assessment of injury risks;
- comment on the strengths and weaknesses of new epidemiologic designs such as the case-crossover design;
- comprehend the basic techniques and uses of qualitative research in order to explore attitudes, identify potential barriers to the adoption of safety measures, or gain the depth of understanding required to develop a programme that will actually motivate changes in human behaviour;
- rank injuries correctly among other causes of ill health with respect to mortality, morbidity, and potential years of life lost;
- distinguish between active and passive safety measures;
- explain why human behaviour is frequently implicated in the causation of accidents, although it may not always represent a priority target for the control of accidents;
- understand the two basic physical equations that allow the determination of injury severity in road traffic accidents, falls, and other injuries caused by transfer of physical energy;
- describe the ten stages that summarize the theory of accident prevention, as conceptualized by William Haddon;
- identify the importance of age and gender, occupation, socio-economic status, and nature of leisure or other activities as risk factors for accidents;
- distinguish between idiosyncratic accident proneness and accident proneness because of the use of alcohol or psychotropic drugs;
- name five transient events that may trigger an accident, and indicate the appropriate methodology for the investigation of the role of such events in accident causation;
- provide examples of how an increase of the area of dispersion and a prolongation of the time required for energy dispersion reduce the severity of injuries;
- explain why the severity of injuries is not a linear function of the speed of a vehicle involved in a crash;
- call upon the simple equations of physics that underlie the role of seat belts and protective helmets in reducing the severity of injuries;

- indicate the approximate fraction of corresponding fatality that could be reduced by the use of seat belts, child car restraints, and helmets;
- identify at least one major difficulty in conducting an epidemiologic study aiming to assess the role of alcohol drinking on accident causation;
- explain why the incidence of home and leisure accidents is grossly underestimated;
- indicate whether active or passive safety measures should be a priority in occupational settings; and
- present how study results can be translated in injury prevention and control.

Teaching contents

Terminology, classification, and sources of data

Issues of terminology, classification, comparability, and data sources need to be presented, even though it takes an extraordinary teacher to make this presentation without boring the students. To this end, instructors can recall their past experience and prepare case studies in order to emphasize points of interest and assist the students to

- define injuries as the consequences of involuntary transfer of energy, as suggested by Haddon (1973);
- distinguish the term 'injury' from 'accident', as the latter term overestimates the role of chance and discourages efforts to identify risk factors for injuries (Langley 1988);
- differentiate unintentional injuries from intentional injuries, namely, injuries caused by violence or which are self-inflicted (including suicides or suicide attempts).

become familiarized with widely used classification systems describing the conditions of injurious events; such systems include the supplementary classification of external causes of injuries and poisonings described in the World Health Organization (WHO) *International Classification of Diseases* (currently in its tenth revision; see <http://www.who.int/classifications/icd/en/>), as well as classification systems used for prevention purposes, such as the Nordic Medico-Statistical Committee (NOMESCO; NOMESCO 1996) and the International Classification of External Causes of Injury (ICECI; see <http://iceci.org/ >). Lastly, scales should be presented which focus on the extent of anatomic damage (e.g. the Abbreviated Injury Scale (AIS), the Injury Severity Score (IBS), and the New Injury Severity Score (NIBS)), focus on the functional severity of

a trauma (e.g. the Revised Trauma Score (RTS) and the Glasgow Coma Scale (GCS)), or combine anatomic location and the functional severity of trauma (e.g. the Paediatric Trauma Score (PTS); Baker et al. 1974).
Students should be reminded that correct and detailed classification is a prerequisite for

- the comparability of data between and within countries;
- the assessment of time trends;
- the evaluation of preventive efforts; and
- the identification of hazardous consumer products.

The comprehensiveness of data on death from accidents is dependent on documentation recorded in death certificates or coronial records. The course instructor can ask the students: 'Do you think that this data is always reliable?' The answer is that it may not have sufficient detail, particularly with respect to the place where the accident occurred, and that there can be over- or under-reporting generated by the misclassification of suicides as accidental deaths, or when there is a long delay between an accident and the subsequent death. Even among the European Union member states, there may be a variable time limit to assign a hospital death to an injury. Thus, when the death occurred, for example, within a week of the accident, this accident is likely to be considered as the underlying cause of death, whereas another adverse condition may be cited as the underlying cause when the death occurred much later after the accident.

In contrast to mortality statistics, morbidity data, including that derived from hospital admissions, are generally of limited reliability, with the exception of that derived from hospital statistics in Scandinavian countries, New Zealand, some states in Australia, and some provinces in Canada, as these areas employ a unique identifying number system which allows data linkage. Even in these areas, however, accidents not requiring hospitalization are inadequately recorded. In many countries, there is now a trend to establish large databases covering all types of injuries that require hospital contact, although not necessarily hospital admission. Such databases are invaluable when the total dimensions of the injury problem need to be evaluated and when hospital statistics are unreliable or of questionable validity (Christoffel and Gallagher 1999).

Descriptive epidemiology

The course instructor should highlight the following points:

1. Injuries represent the third most common cause of death in the developed countries, following diseases of the circulatory system and malignant neoplasms.

2. Injuries are responsible for as many or more years of life lost as cardiovascular disease and malignant neoplasms are, because the latter two categories generally affect older people, whereas injuries are concentrated among younger population groups.
3. Road traffic injuries are the eighth leading cause of death globally and the leading cause of death for young people aged 15–29. Current trends suggest that by 2030 road traffic deaths will become the fifth leading cause of death unless urgent action is taken. Eighty-eight countries have reduced the number of deaths on their roads but the total number of road traffic deaths remains unacceptably high at 1.24 million per year (WHO 2013).
4. Mortality rates from accidents are twice as high among men compared to women, and mortality by age follows a typical pattern (Baker 1992).
5. In proportional terms, more than 50 per cent of fatal injuries are caused by road traffic accidents, about one-third from home and leisure accidents, and 10–15 per cent are occupational in nature.
6. Fatal accidents represent the peak of the pyramid that depicts the magnitude of the injury problem. Every case of fatal injury corresponds to about 30 cases of serious injury that require hospitalization and may lead to permanent disability, and about 500 cases of injuries that require some form of medical attention.
7. In developing countries, injuries account for approximately 9 per cent of mortality and morbidity, with the majority of injuries due to violence (self-inflicted and interpersonal) and road traffic injury.

Etiologic considerations

Injuries are caused when energy is transferred to the human body at a rate and density that exceeds the resistance of human tissues (Robertson 1992). Human behaviour, however, can be a critical determinant of the frequency and severity of injuries, by affecting the frequency and the duration of exposure to various sources of energy as well as the rate and density of energy transfer in susceptible individuals. Indeed, most accidents have multifactorial etiology, and their prevention or the reduction of severity of the resulting injuries can be accomplished through intervention at either the medium by which energy is transferred or by changing the behaviour of the susceptible individual and thus reducing the frequency or duration of exposure to the source of energy. The course instructor can refer to the serious consequences for car occupants following a crash and prompt a discussion on how such consequences can be avoided through a change in the behaviour of the driver (so that the probability

of an accident is reduced) or through passive protection measures built into the car and aimed at ameliorating the consequences of the accident.

The multifactorial etiology of accidents is not conceptually different from that of most diseases. It is instructive to point out the existing parallels, for example, the etiologic sequence of an agent (be that *Plasmodium vivax* in the case of malaria, or kinetic energy in the case of a fracture) that via a vector (be that a mosquito or a motor vehicle, respectively) and following an exposure event (be that a mosquito bite or a car crash, respectively) results in human damage (malaria or a fracture, respectively). In all instances, prevention can be accomplished by focusing on the etiologically interacting factor that appears more amenable to some form of intervention. A prime example that can be used by the instructor is childhood poisoning with aspirin, because of the number of factors acting in conjunction. The normal curiosity of babies, prompting them to taste everything they touch; the accessibility of pills when they are not packaged in childproof containers; the appealing colour of children's aspirin; the large number of pills in each package; and careless parental behaviour (e.g. leaving medicines and other dangerous chemicals in easily accessible places, and lack of supervision) are in this case the constellation of responsible factors that can be depicted by a triangle: the poisoning agent (aspirin), the host (the child), and the environment (e.g. container, storage, supervision, etc.).

In that instance, it is difficult to allocate degrees of responsibility to individual factors, even when the other interactive causal factors remain constant. Therefore, the primary objective of epidemiologic research towards the prevention of injuries relies on the recognition of the multifactorial origin of injuries and the identification of the 'weak link' in the causal sequence. The use of safe packaging that precludes easy opening by toddlers can solve the problem to a very large extent without attempting to change behavioural interacting factors that are known to be difficult to modify. This is why there is a general tendency to implement passive prevention measures, which are essentially technological, rather than active prevention measures, which focus on behavioural modification. Examples of passive prevention measures are the use of childproof packaging for medicines and chemicals, the use of balcony fences that are sufficiently high and dense to restrict toddlers from falling from a height, the installation of smoke alarms, and the regulation of thermostats so that hot water burns can be avoided. It should be emphasized, however, that most accidents involve man-made factors, whether in connection with motor vehicles, highways, poisonous chemicals, consumer products, or swimming pools. In fact, the environment that causes injuries is the one that has been created by mankind; in this sense, it can also be redesigned to be safer (Christoffel and Gallagher 1999); for example, devices such as the nasal continuous positive airway pressure pump, used by

patients suffering obstructive sleep apnoea syndrome, are effective in preventing road traffic accidents (Antonopoulos et al. 2011).

It is also important to understand that the recognition of the multifactorial origin of accidents and subsequent injuries relies on accurate and precise measures of risk. Considerable efforts are being undertaken to improve the quantity and quality of data used in accident prevention research; consequently, analyses have become increasingly complex and insightful, enabling the identification of important associations not possible by analysis of numerator-based data alone. One such improvement, albeit nascent, is the measurement of episodic person-time exposure. For example, Ballesteros and Dischinger (2002) found that, among young drivers in the US, the crash rate per driver, unadjusted for exposure, was highest for 16-year-olds (155.7 crashes per 1,000 drivers) and lowest for 20-year-olds (85.2 crashes per 1,000 drivers), a rate ratio for the two age groups of 1.8. However, when the rates were adjusted for exposure, namely, vehicle miles driven, the rate ratio increased nearly fivefold, to 8.4. This finding suggests that experience, rather than age, may be more important in predicting the probability of motor vehicle crashes. Clearly, understanding person-time exposure can improve our understanding of risk.

A qualitative approach towards injury epidemiology

Quantitative research remains the primary and fundamental path for acquiring data on both the causation and prevention of injuries. Given that injury epidemiology is a multidisciplinary field and that injury prevention relies both on passive prevention measures and human behaviour, health professionals and practitioners are expected to have a basic understanding of the concept and applications of qualitative research if they are to practice injury prevention. Thus, the course instructor should attempt not only to convey to the students the knowledge and skills required to distinguish between quantitative and qualitative techniques but also how to apply the appropriate methodology for the aims of their study. An example of a short syllabus follows.

1. 'What techniques are used for collecting data in qualitative research?' Mainly, these techniques comprise in-depth interviews, focus groups, observation, case studies, and life story and biographical approaches. These designs have traditionally been employed in the social sciences and relatively recently they have also been included in the research agendas of health-care professionals and injury prevention investigators.
2. 'What is the difference between qualitative and quantitative research?' In contrast to the assessment of a relation between exposure and outcome and the quantification of effect estimates, qualitative approaches as a rule are

aimed at exploring 'research questions' about peoples' experiences and perceptions, to come up with the essence of meanings and understand how particular phenomena are generated in particular contexts. Thus, qualitative studies seek answers to questions about the 'what', 'how', and 'why' of a phenomenon, whereas quantitative studies seek answers to questions about 'how many' or 'how much'. If the study objective is, for example, to explore what the attitudes of adolescents towards helmet use are, it would be more appropriate to use a qualitative method to elicit in-depth understanding of the reasons. If, on the contrary, the study objective is to assess how many adolescents wear a helmet while riding a motorcycle, a quantitative method would be more suitable. In other words, the basic goal of qualitative studies is to explore and to comprehend profoundly a phenomenon, rather than to measure it (Green and Thorogood 2006).

3. 'How is data analysed in qualitative research?' Qualitative and quantitative approaches also clearly differ in terms of how the data are analysed. Quantitative research requires the reduction of phenomena to numerical values in order to carry out statistical analyses. By contrast, qualitative research involves collecting data in the form of naturalistic verbal reports, and the analysis conducted on this is textual. Qualitative studies typically analyse words, not numerical data, and the main purpose of such analysis is to identify patterns and 'themes' across language and within a range of theoretical frameworks, from essentialist to constructionist and even discourse analysis (Smith 2008).
4. 'What are the strengths and limitations of the qualitative approach?' Qualitative research is, by definition, stronger on long descriptive narratives than on statistical tables. Specifically, qualitative methods can elicit rich, contextual data and provide valuable insights into the meanings of decisions and actions; by contrast, the small study samples and flexible design of qualitative research limit significantly the generalizability of the results (Ulin et al. 2005). Moreover, both the involvement of respondents, not merely as subjects but as participants who engage in an active process, and the subsequent interaction between them and the researcher, or even among themselves (i.e. in the case of a focus group study), could endanger the validity and reliability of the study, in quantitative research terms. However, under the qualitative research perspective this approach is considered to be a valuable methodological tool, and both the personal involvement in the communication with the participant and the perceived subjectivity of the researcher are considered necessary prerequisites for acquiring rich data and producing insightful data analysis (Smith 2008).
5. 'Can qualitative studies be considered the opposite of quantitative research?' The course instructor should emphasize that not only are these approaches

not opposing but in fact they can be used to supplement each other. Different aspects of this complimentary use include (a) starting with a qualitative study in order to explore a particular topic and generate a working hypothesis and then setting up a quantitative study, (b) initiating a quantitative study in order to establish a sample of respondents and/or the broad contour of the field and then using a qualitative study to explore a specific issue with the sample of respondents identified from the quantitative study, and (c) engaging in a qualitative study that uses quantitative data in order to define the results in a broader context.

6. 'How should the qualitative research approach be taught to injury epidemiology students?' A primary challenge of teaching traditional widespread qualitative methods of analysis such as interpretative phenomenological analysis and thematic analysis to students new to qualitative research stems from the frequent domination of quantitative methods in the field. Indeed, most students believe that the quantitative approach is the only 'scientific' way to do research. When we introduce students to the assumptions and values of qualitative research, we have to make clear that subjectivity and interaction between the researcher and research participants do not produce bias that undermine the research but instead are essential components of good qualitative research practice. 'Learning by doing' appears to be the most appropriate way to teach qualitative research. Understanding the application of the method in practice seems to allow space for the theory to make sense.

In addition to learning theory through practice, there are a number of key strategies that can be employed in teaching qualitative analytic methods, including (Braun and Clarke 2013)

- using real, primary data (ask students to generate their own data after determining a research topic and question, designing a data collection tool and all relevant research materials, and obtaining ethical approval);
- illustrating concepts and steps with examples from previous studies, to show students what codes, thematic maps, and coded extracts look like;
- using excellent published studies as examples;
- using a reflexivity exercise—namely, ask students to reflect and note their assumptions and ideas about the research topic;
- demonstrating the importance of coding by using a structured coding exercise;
- using out-of-classroom exercises to expedite classroom learning; and
- demonstrating 'bad' practice as it arises.

Complete the teaching with a discussion of theory; namely, discuss the importance of engaging with theory issues such as whether we are working bottom-up from the data or reading the data through the lens of existing theory. In turn, the course evaluation could be based on this plan by asking students to work on their own qualitative research projects, starting from the construction of the research question, the collection of data, content analysis, thematic analysis, and ending with the write-up of their original qualitative study in injury epidemiology.

The theory of accident prevention

In presenting the theory of accident prevention, trivializing the subject by invoking a collection of simple unrelated advisory statements should be avoided, however useful these statements may be. Instead, the teacher should discuss the prevention of accidents in the context of the ten discrete countermeasure strategies, as conceptualized by the late William Haddon, a pioneer in this field (Haddon 1970), and presented in the Haddon matrix (Table 22.1). To this end, various techniques can be employed to reduce accidents, including preventing the build-up of energy, reducing the initial amount of energy, preventing the release of energy, carefully controlling the release of energy, and separating the energy being released from the living or non-living object.

At this stage, it is worth repeating to the students that the transfer of energy causes injuries to susceptible individuals and noting that the transfer of mechanical energy is the most common mechanism of injuries. Indeed, the transfer of mechanical energy causes both falls and road traffic accidents. Students are

Table 22.1 The Haddon matrix

	Human factor	**Vector/product**	**Physical environment**	**Socio-economic environment**
Pre-event	Is there a predisposition? Is there overexposure to risk?	Is the vector hazardous?	Is the environment hazardous? Is it possible to reduce hazards?	Does this encourage or discourage risk-taking and hazard?
Event	Able to tolerate force/energy transfer?	Does the vector provide protection?	Does environment contribute to injury during event?	Does the environment contribute to injury during event?
Post-event	How severe is the trauma?	Does the vector contribute to the trauma?	Does the environment add to the trauma after the event?	Does the environment contribute to recovery?

usually eager to learn how to quantify the expected energy transfer and thus estimate the expected damage. Students should be reminded of two basic equations in physics: $E = \frac{1}{2}mv^2$, and $v = \sqrt{2}\,gs$, where E = energy; m = mass; v = velocity; g = acceleration or deceleration; and s = the distance covered under the influence of acceleration or deceleration.

As a result, two cars crashing while moving towards each other at a relatively slow speed of 25 kilometres/hour (~7 metres/second) each confer to their occupants consequences equal to those that would have been created if the occupants were falling from a height of 10 metres (under the influence of gravity, the reference speed should be 25 kilometres/hour × 2 = 14 metres/second). Solving the second of the two equations for g indicates that the deceleration, which reduces the speed to zero, is inversely proportional to the distance covered by an individual during that crash. This is the principle underlying the widely used passive safety measure that consists of the gradual (prolonged) collapse of the anterior or posterior part of the car and which allows the passenger cabin to remain intact.

In a similar spirit, in explaining the risk factors for electrocution, the teacher should invoke the principles of physics regarding electricity. The consequences of electrocution are a function of the intensity of the electric current, which in turn is inversely related to the electric resistance of the body, which is substantially reduced (more than tenfold) when the body is wet.

Risk factors for accidents

There are no risk factors that are important for every type of accident in every population group. In general, however, the frequency and severity of accidents depend on a series of factors such as

- age and gender (men are usually at higher risk, and the risk is higher among the 15–24 year age group and towards the end of life);
- occupation type (construction workers are generally at higher risk) or type of leisure activity (the level of risk is considerably higher among rock climbers);
- the safety coefficient of structures, processes, consumer products, etc., which is a function of legislation, regulations, and safety standards;
- the area of dispersion and the prolongation of the time required for energy dispersion, which are inversely associated with the risk and the severity of accidents (e.g. helmets increase the dispersion of the energy from a crash, whereas seat belts increase both the time for deceleration as well as the spread of the energy over the various parts of the body);
- accident proneness, which may be caused by neurologic diseases or created through the use of psychotropic drugs or by some behavioural patterns

(personality and idiosyncratic characteristics may also indicate a tendency towards risk behaviours and accident proneness);

- transient high-risk behaviours (e.g. emotional stress, attention detraction via using a cellular phone or changing a radio station, etc.), which dominate the etiology of about two-thirds of accidents (Petridou et al. 1998);
- genetic and environmental influences, and the exposure to accidental life events (Distel et al. 2011); and
- sleep-disordered breathing (a relationship between sleep-disordered breathing and depression has been proposed, and screening and quantification of sleep-disordered breathing should be considered in depression research; Cheng et al. 2013).

Although behavioural problems contribute to the majority of accidents, it is not easy to identify individuals at high risk of such behaviours on the basis of a psychological profile. Instead, high-risk individuals should be identified through their tendency to drink or to use psychotropic drugs while driving, or among those who drive very long distances, like truck drivers. The teacher should handle this issue carefully to avoid the stigmatization of a category of individuals who are over-represented in injury statistics. This part of the course also provides an opportunity for the teacher to discuss the merits of targeting high-risk populations versus a whole population approach. It is important to stress the fact that widespread behaviours can be instrumental in the causation of several accidents. At this point, it is also possible to discuss the philosophy behind legal measures that have a strict paternalistic dimension, an issue that has been controversial in libertarian circles: 'It is not the business of the state to force me to use a seat belt, so long as I do not endanger the life of anybody than myself' is a hollow but commonly used argument.

It has always been a challenge for investigators to develop formulas that would derive the probabilities of accidents (risks) associated with various occupational or leisure activities while taking into account exposure-related variables. Such formulas are desired by professional organizations as well as insurance companies, although the process of determining the formulas may be considered a distraction for those investigators more interested in findings with a greater potential for the prevention of injury itself.

Major accident categories

Road traffic accidents

Road traffic accidents (RTAs) are responsible for about 50 per cent of accidents in most developed countries, and a gradually increasing proportion in

developing countries. Thus, they deserve special consideration. The teacher, however, should be careful to avoid sentimentalism and instead concentrate on issues that emerge from the examination of the basic issues related to the causation of injuries. In an RTA, the severity of injuries depends on the deceleration imposed on the vehicle occupants during the so-called 'secondary' or 'human' stage of the crash. During this stage, the inertial speed of the occupants is suddenly reduced to zero as their bodies crash into the immobilized vehicle. Similarly, the speed acquired by the pedestrians after a vehicle hits them is suddenly reduced to zero as they crash into the road or another immovable object.

Deceleration is a function of the ratio v^2/s, where v = velocity and s = the distance covered by the body during the time that the speed is reduced to zero during the secondary crash. The numerator demonstrates why high speeds have disproportionately large consequences, since a crash at 60 kilometres/hour has consequences that are almost ten times more serious than a crash with a speed of 20 kilometres/hour. The course instructor should then try to demonstrate the importance of the following two safety measures.

1. Seat belt use. During the crash of the vehicle (primary crash), the anterior mechanical parts of most modern vehicles are squeezed in such a progressive way that the energy is gradually absorbed. As a consequence, the reduction of the speed of the vehicle to zero is accomplished in a distance of about 30 to 50 centimetres. The car cabin and the belted occupants follow the same trajectory. In contrast, a passenger who is not using a seat belt crashes (because of inertia), with the full original speed of the vehicle, into the anterior part of the car. In this instance, the reduction of the original speed to zero takes place within a distance of only a few centimetres. The consequences of the high deceleration of the body should be obvious. It is clear that seat belts are useful because they force the users to follow the movement of the vehicle and thus allow them to exploit the much smaller deceleration of the primary crash rather than the high deceleration that would have characterized the secondary crash. A seat belt has been estimated to reduce the probability of serious injury or death by about 40 per cent.
2. Availability of functioning airbags. The expansion of the airbag at the moment of crash also contributes to the prolongation of the distance covered as velocity declines to zero. Moreover, the airbag conveys additional protection because the released energy is spread over a larger area. It should be obvious that seat belts, preferably three-point seat belts, and airbags (in passenger cars) should be available for all car occupants.

The teacher should also make sure that sufficient information is provided to the students about

- the significant additional protective effect imparted for children riding in the rear seat of a vehicle by the use of a child car restraint—in fact, the effectiveness of child car restraints in reducing serious injuries and deaths has been estimated to exceed 70 per cent;
- the magnitude of protection imparted by the use of suitable and properly functioning motorcycle helmets, with the probability depending on the speed of the motorcycle involved—indeed, helmet use can reduce the risk and severity of head injuries by about 72 per cent, and the likelihood of death by up to 39 per cent; and
- the substantially advanced prevention of RTAs by the investigation of place and time variables, namely, place clustering, time periodicity, and time–place clustering.

It is unfortunate that deductive epidemiology studies are not more frequently used for road traffic injuries, because small 'outbreaks' of injuries may be as common as outbreaks of infectious diseases. Nowadays, in a number of countries, 'black-spot programmes' are in place, whereby sites on the road network in which multiple RTAs have occurred are subjected to intense assessments (road audits) in order to design safety into the road environment.

The emphasis on injury epidemiology in relation to behavioural factors should not lead the instructor to underestimate the critical importance of technical factors. These factors refer to vehicles or the road system and may focus on secondary prevention (e.g. the use of large cars, of seat belts, or of unyielding car cabins) as well as on primary prevention (e.g. increased visibility, anti-blocking systems, or elevated braking lights). Research and implementation of these developments by the contemporary car industry, as well as road auditing, have a potentially high contribution to the reduction of the RTA toll.

A final but epidemiologically complicated issue is the role of ethanol and recreational drugs in the causation of road traffic injuries. None has the right to challenge the basic principle that drinking and driving should be separate activities. Nevertheless, one should also admit that the undertaking of analytic studies to evaluate the relationship between drinking of alcoholic beverages or consumption of recreational drugs and driving, especially among the younger or relatively inexperienced drivers, is methodologically challenging.

Suicide

As suicide is the leading cause of injury-related death in both high-income and low-to-middle-income countries, the teacher cannot abstain from discussing this topic, despite the fact that this chapter does not essentially target intentional injuries.

- Globally, over the past four decades, the rate of suicide has remained stable, with the rate of suicide three times higher for men (17.6 per 100,000 population) than for women (5.6 per 100,000 population; Australian Institute for Suicide Research and Prevention 2003).
- Rates of suicide are high in Eastern Europe, moderate in Western Europe, North America, and the Pacific, and low in Latin America, some South East Asian countries, and the Middle East.
- Following the increasing European trends, the rate of suicides and self-harm behaviours among adolescents and youths has increased over 24 per cent from 2007 to 2009. In particular, the Greek Ministry of Health recently reported an increase of 40 per cent in the overall suicide rate, a fact that indicates the role of social and cultural factors such as the economic crisis in the increase of suicides.
- Underlying causes of suicide include depression, which is associated with sunlight exposure (Papadopoulos et al. 2005), substance abuse, particularly alcohol, and unemployment (Australian Institute for Suicide Research and Prevention 2003). The identification and treatment of mental illness has been the main strategy for suicide prevention, although it should be noted that there are few strategies reported to be highly effective in reducing the rates of suicide. The teacher needs to emphasize the need for the highest level of evidence (namely, from randomized controlled trials), in order to assess the efficacy of strategies for suicide prevention.

Home and leisure accidents

Home and leisure accidents are extremely common but their low fatality reduces their importance in terms of mortality rates.

- A characteristic of home and leisure accidents and of their associated injuries is that their epidemiologic study relies on very limited data in comparison to those available for the study of RTAs.
- Until recently, progress in the prevention of home and leisure accidents has relied on descriptive information from case series or even case reports.
- Examples of preventive measures that were instituted on the basis of simple descriptive epidemiologic data are usually given by the mechanism of injury; that is, falls and exposure to inanimate mechanical forces, burns, poisoning, drowning and submersion, and contact with heat or hot substances. Examples of such preventive measures can be found in textbooks concerning the epidemiology and prevention of injuries (Baker 1992; Christoffel and Gallagher 1999).

Occupational injuries

Epidemiologic studies on occupational injuries, however imperfect, have converged in indicating that the prevention of these injuries should rely on passive measures and not require special attention or precautions on the part of the worker. The role of epidemiology, in this instance, is to document the existence of excess risk in various types of jobs or activities and to provide risk profiles and attributable risks. The epidemiologic approach does not differ much from that commonly used in occupational medicine (i.e. calculation of standardized ratios or, at a minimum, proportional ratios).

Upgrading injury epidemiology

In developed countries, injuries represent the most common cause of death during the first four decades of life and reduce life expectancy as much as malignant neoplasms and cardiovascular disease. Likewise, RTAs account for more years of potential life lost than any other specific nosologic entity. Perhaps more importantly, deaths from injuries are at least twenty times more frequent than deaths from AIDS, environmental pollution, passive smoking, ozone depletion, and many other causes that have galvanized social reaction. It is generally accepted that advancements in epidemiology have preceded the curtailment of mortality rates from infectious and cardiovascular diseases and several forms of cancer. It is clear, however, that the epidemic of injuries in many populations remains unabated and there is an urgent need to expand the inductive and deductive epidemiologic investigation of injuries.

Recent advancements in the treatment of traumatic injuries have resulted in significant gains concerning the avoidable fraction of mortality and morbidity due to injuries. For example, based on WHO data and a systematic literature review, it has been estimated that each year worldwide there are approximately 400,000 in-hospital trauma deaths due to traumatic bleeding. If patients received tranexamic acid (TXA) within one hour of injury, approximately 128,000 deaths might be averted, whereas if patients received TXA within three hours of injury, approximately 112,000 deaths might be averted (Ker et al. 2012). On the other hand, methods for evaluating knowledge-transfer and exchange interventions will offer great benefits in the new epidemiological research era (see Kramer et al. 2013 for a description of such a method, which is based on a synthesis of the theoretical frameworks underpinning (1) the promoting action on research implementation of health services model, (2) the transtheoretical model of change, and (3) a model of knowledge use). In addition, interdisciplinary studies in epidemiology, sociology, and social psychology may further contribute to the understanding of the cultural, socio-economic,

and psychosocial aspects of injury, since culture and the modern way of life both are produced in and reproduce an environment of ambiguity about risk and safety related to the figured consequences of modernity and the challenges of the postmodern 'risk society' (Beck 1998).

Course evaluation and students' assessment

For lower-level students, a demonstration of the ability to recall basic data of general applicability, especially for study design, and the ability to link particular preventive measures to Haddon stages and basic laws of physics should represent adequate evidence that they have met the course objectives. A series of multiple-choice questions distributed to students for self-evaluation and subsequently discussed in the class provides an opportunity to refresh the recently acquired knowledge. Short essays or multiple-choice questions could assist in the overall assessment of the students at the end of the course.

Advanced-level students can be evaluated via the submission of a study design to investigate the role of a particular condition or activity in the causation of accidents. Alternatively, students can be involved in the preparation of a case study and describe issues associated with the assessment of an intervention project. In particular, for qualitative research methods, students may be asked to work on their own qualitative research projects while being assessed step by step in all research phases, particularly with respect to interviewing, content analysis, and the write-up of their original study. The process will be challenging for the student and, in some cases, even for the instructor, as injury epidemiology is full of subtle and complex issues.

References

Antonopoulos, C. N., Sergentanis, T. N., Daskalopoulou, S. S., and Petridou, E. T. (2011) Nasal continuous positive airway pressure (nCPAP) treatment for obstructive sleep apnea, road traffic accidents and driving simulator performance: a meta-analysis. *Sleep Medicine Reviews*, **15**: 301–10.

Australian Institute for Suicide Research and Prevention. (2003) *International Suicide Rates: Recent Trends and Implications for Australia*. Canberra: Australian Government Department of Health and Ageing.

Baker, S. P., O'Neill, B., Haddon, W. Jr, and Long, W. B. (1974) The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma*, **14**: 187–96.

Baker, S. P., O'Neill, B., and Karpf, R. S. (1992) *The Injury Fact Book*. New York: Oxford University Press.

Ballesteros, M. F. and Dischinger, P. C. (2002) Characteristics of traffic crashes in Maryland (1996–8): differences among the youngest drivers. *Accident; Analysis and Prevention*, **34**: 279–84.

Beck, U. (1998) *World Risk Society*. Cambridge: Polity Press.

Braun, V. and Clarke, V. (2013) Teaching thematic analysis. *The Psychologist*, **26**: 120–3.

Cheng, P., M, D. C., Chen, C. F., Hoffmann, R. F., Armitage, R., and Deldin, P. J. (2013) Sleep-disordered breathing in major depressive disorder. *Journal of Sleep Research*, **22**: 459–62.

Christoffel, T. and Gallagher, S. S. (1999) *Injury Prevention and Public Health: Practical Knowledge, Skills and Strategies*. Gaithersburg, MD: Aspen Publishers, Inc.

Distel, M. A., Middeldorp, C. M., Trull, T. J., Derom, C. A., Willemsen, G., and Boomsma, D. I. (2011) Life events and borderline personality features: the influence of gene-environment interaction and gene-environment correlation. *Psychological Medicine*, **41**: 849–60.

Green, J. and Thorogood, N. (2006) *Qualitative Methods for Health Research*. London: Sage Publications, Ltd.

Haddon, W. Jr. (1970) On the escape of tigers: an ecologic note. *American Journal of Public Health and the Nation's Health*, **60**: 2229–34.

Haddon, W. Jr. (1973) Energy damage and the ten countermeasure strategies. *Journal of Trauma*, **13**: 321–31.

Ker, K., Kiriya, J., Perel, P., Edwards, P., Shakur, H., and Roberts, I. (2012) Avoidable mortality from giving tranexamic acid to bleeding trauma patients: an estimation based on WHO mortality data, a systematic literature review and data from the CRASH-2 trial. *BMC Emergency Medicine*, **12**: 3.

Kramer, D. M., Wells, R. P., Carlan, N., Aversa, T., Bigelow, P. P., Dixon, S. M., and McMillan, K. (2013) Did you have an impact? A theory-based method for planning and evaluating knowledge-transfer and exchange activities in occupational health and safety. *International Journal of Occupational Safety and Ergonomics*, **19**: 41–62.

Langley, J. D. (1988) The need to discontinue the use of the term 'accident' when referring to unintentional injury events. *A*ccident; Analysis and Prevention, **20**: 1–8.

Nordic Medico-Statistical Committee (NOMESCO). (1996) *NOMESCO Classification of External Causes of Injuries*. Copenhagen: Nordic Medico-Statistical Committee.

Papadopoulos, F. C., Frangakis, C. E., Skalkidou, A., Petridou, E., Stevens, R. G., and Trichopoulos, D. (2005) Exploring lag and duration effect of sunshine in triggering suicide. *Journal of Affective Disorders*, **88**: 287–97.

Petridou, E. (1997) Epidemiology and injury prevention. *Injury Prevention*, **3**: 75–6.

Petridou, E., Mittleman, M. A., Trohanis, D., Dessypris, N., Karpathios, T., and Trichopoulos, D. (1998) Transient exposures and the risk of childhood injury: a case-crossover study in Greece. *Epidemiology*, **9**: 622–5.

Robertson, L. S. (1992) *Injury Epidemiology*. New York: Oxford University Press.

Smith, J. A. (2008) *Qualitative Psychology: A Practical Guide to Research Methods*. London: Sage.

Ulin, P. R., Robinson, E. T., and Tolley, E. E. (2005) *Qualitative Methods in Public Health: A Field Guide for Applied Research*. San Francisco: Jossey-Bass.

WHO. (2013) *Global Status Report on Road Safety 2013: Supporting a Decade of Action*. Geneva: WHO.

Chapter 23

Dental epidemiology

Georgios Tsakos and Vibeke Baelum

Introduction to dental epidemiology

Most dental students are introduced to their future profession through courses that reward reproduction of factual information on topics such as anatomy, dental materials, and biochemistry. Later in the curriculum, they undergo comprehensive clinical training with the aim of acquiring technical expertise and becoming precision mechanics. These characteristics of their training, with an emphasis on technical details, tend to shape dental students more as passive learners rather than as inquisitive, active learners who adopt the scientific paradigm in medicine and dentistry. In addition, these features also tend to affect the views of the students on dentistry and on their future role as carers for the individual or even specific parts of the oral cavity. However, epidemiology is concerned with the study of health and disease in populations, where the group—not the individual—is the focus of interest. Unlike the biological and clinical sciences, epidemiology is a discipline based on logical deduction and scientific reasoning. It is therefore not surprising that dental students tend to find epidemiology more difficult than the core clinical topics, just as many consider epidemiology and biostatistics as topics at the fringe of their future role as practising dentists.

The overall aim is to provide the students with the necessary theory and tools to become true professionals, that is, critical and theoretically informed individuals within their particular field. Ideally, we would like to use teaching and learning strategies that combine the best of traditional modes with more recent approaches that shift the emphasis from the teacher to the student, hence providing a meaningful context for learning and promoting self-awareness of the learner, with the teacher facilitating (rather than directing) the whole process. The long-term aim is deeply rooted in a belief that contemporary professionals need the skills of lifelong learning and the ability to deal with uncertainty and value conflicts. However, we acknowledge that this approach is best suited for active learners and seems less compatible with the current reality of predominantly passive learning in the undergraduate dental curricula; however, it can

be easier tailored to postgraduate degrees in dental disciplines, particularly those that have a strong epidemiological element (e.g. MSc courses in dental public health). Undergraduate dental students are primarily considered *consumers of research* whereas postgraduate students are also potential or actual *producers of research*. This leads us to focus primarily on clinical epidemiology for the undergraduate dental student level.

For both undergraduate and postgraduate teaching, we would like to stress the importance of a comprehensive study guide that will assist students in their learning. By specifying learning outcomes, a good study guide indicates what should be learnt and helps students to set their own objectives and plan their learning. It should contain the answers to the questions 'Where are we?', 'How did we get here?', as well as 'Where are we going now?' and 'How will we get there?' Thus, it provides guidance like a good tutor but without the need for excessive staff–student contact. We strive to design our teaching on the principle of alignment of (1) the curriculum taught, (2) the teaching methods used, (3) the assessment procedures used, (4) the climate created in the interaction with students, and (5) the institutional climate, rules, and procedures that must be followed (Biggs 2003); however, we often have to come to terms with the fact that this can be quite difficult to achieve.

Teaching objectives

The objective is to introduce dental students to the basic principles and methods of epidemiology, with a focus on clinical epidemiology as it relates to dental epidemiology in particular. As we consider the undergraduate and postgraduate dental students as *consumers of research*, we aim at producing *critical consumers* (Greenhalgh 1998). Therefore, the whole course is influenced by the concepts of evidence-based practice. And, as epidemiology is the most relevant discipline to health-care decision-making (Muir Gray 1997), its teaching should have a critical and central role in the whole curriculum. At the undergraduate level, epidemiology is usually taught in combination with a basic course in biostatistics (Table 23.1), where the students are taught the most frequently used statistical methods in the dental and medical literature (Kirkwood and Sterne 2003).

The biostatistical part consists of lectures and exercises based on textbooks and on data from dental articles or relevant studies. A textbook in biostatistics covers the curriculum and the examination requirements (Kirkwood and Sterne 2003). For the undergraduate teaching of epidemiology, there are several useful introductory epidemiology textbooks (e.g. Bonita et al. 2006) which cover the essentials of epidemiology (e.g. Coggon et al. 2003; Greenberg et al.

Table 23.1 Student learning in different aspects of statistics and epidemiology

Category	Parameters and tests	Level of learning
Exploring, summarizing and presenting data (descriptive statistics)	Mean, median, mode, Gaussian distribution, standard deviation, percentiles, etc.	Understand, master, and interpret
	Graphs	Understand and interpret
Making inferences from data	Confidence intervals and hypothesis testing	Understand, master, and interpret
Parametric statistics	*z*-test, *t*-test, Pearson's correlation coefficient	Understand, master, and interpret
	ANOVA	Understand and interpret
Non-parametric statistics	Mann–Whitney rank sum test, Wilcoxon signed test, chi-square test, Spearman's correlation coefficient	Understand, master, and interpret
Multivariable methods	Mantel-Haenszel's summary odds ratio and chi-square test	Understand, master, and interpret
	Linear regression	Understand and interpret
	Logistic regression	Understand and interpret
Survival analysis	Life tables (actuarial analysis) and Kaplan–Meier product limit estimates	Understand and interpret
Measures of disease frequency	Prevalence, incidence rate, cumulative incidence	Understand, master, and interpret
Measures of association	Incidence rate ratio, relative risk, odds ratio	Understand, master, and interpret
Standardization	Direct and indirect standardization	Understand and interpret
The diagnostic matrix	Sensitivity, specificity, positive and negative predictive values	Understand, master, and interpret
Intra- and extra-examiner variability	Percent agreement, Bland–Altman plots	Understand and interpret
	Kappa	Understand, master, and interpret

2004), while there are more advanced textbooks may also be considered (Hennekens et al. 1987; Rothman 2002), although they are really more suitable at postgraduate level.

The students are introduced to epidemiology as soon as they have understood and mastered the most elementary statistical methods. As the aim is to integrate statistics into the learning of epidemiology rather than to teach statistics as a separate subject, the teaching responsibility should be shared by teachers from or familiar with both disciplines. The key issues of teaching in

epidemiology are presented in Table 23.2. Generally, the focus is on oral health but during plenary teaching the students are also exposed to key areas from medical epidemiology. The emphasis on validity and reliability in the initial sessions lays the foundations for the discussion of bias and confounding, and these are interlinked with the sessions on study designs in epidemiology. Furthermore, critical appraisal of literature is emphasized and practised, where the

Table 23.2 Teaching dental epidemiology to undergraduates: key issues

Issue	Level of learning
Samples and sampling strategies	Understand and interpret
Validity and reliability	Understand and interpret
Bias and confounding	Understand and interpret
Misclassification	Understand and interpret
Diagnostic tests, screening, the diagnostic matrix	Understand, master, and interpret
Randomized controlled trials	Understand and interpret
The four basic epidemiologic study types:	Understand and interpret
cross-sectional studies	
case-control studies	
cohort studies	
intervention studies—randomized control trials	
Measures of disease frequency and measures of association	Understand, master, and interpret
Causation	Understand and interpret
Confounding	Understand and interpret
Control of confounding in design and analysis	Understand and interpret
Chi-square test, Mantel–Haenszel's test, control of confounding	Understand, master, and interpret
Multivariable analysis in epidemiology:	Understand and interpret
linear regression	
logistic regression	
Oral and dental epidemiology:	Learn
national and international figures and issues	
Descriptive and analytic issues particular to dental epidemiology	Understand
Interpretation of epidemiologic literature	Understand
How to read a paper	Master

extensive use of epidemiological knowledge is highlighted and the hierarchy of evidence is discussed. Indeed, the principles and practice of evidence-based dentistry underpin the whole teaching of dental epidemiology. Proficiency in epidemiology and medical statistics is considered as a core element of *community oral health* (or *dental public health*; the terminology varies across countries and schools but both terms refer in essence to the same discipline).

A proper understanding of the basic theoretical aspects of biostatistics and epidemiology is important. For example, although the students should learn to compute confidence intervals of point estimates, this is usually a trivial exercise when the proper equation has been identified. The conceptual aspect, together with proper understanding and interpretation of a given confidence interval, is far more important. Formulas, algebra, and calculations are thus merely viewed as tools that can be readily accessed once an informed decision has been made about the choice of the appropriate test for a particular problem. All relevant information should be placed on an intranet page made available to the students. Consequently, the focus should be on understanding the uses and limitations of the different statistical methods rather than on accurate calculus.

Content of teaching

There are many oral diseases and conditions, but the two most common are dental caries and periodontal disease. Unlike most medical conditions, both these oral diseases are ubiquitous in the sense that virtually all members of the population show at least mild signs of the diseases. In addition to the concepts of prevalence and incidence, these two diseases therefore call for the use of extent and severity measures to describe the number of teeth affected and depth of lesions. However, the teaching of epidemiology should also cover other important oral health conditions such as oral cancer, malocclusions, and dental trauma, which may provide helpful examples for the understanding of key concepts and epidemiological study designs. In addition, such conditions can also be used to discuss related public health aspects, such as determining what would be considered an important public health problem (Sheiham and Watt 2003; Daly et al. 2013).

Dental caries

The diagnosis of dental caries illustrates a number of problems common to many diseases. These include defining the characteristics of a disease and evaluating diagnostic criteria and diagnostic methods, including diagnostic test validity, and inter- and intra-observer variation.

Dental caries occurrence can be quantified in several ways. Prevalence estimates have typically been of limited usefulness because they tend to be very high. Under such circumstances, the number of affected teeth or tooth surfaces per person provides more detailed information since it reflects differences in the extent of disease. In dental epidemiology it is therefore common to use disease indices that capture this dimension, and dental caries in the permanent dentition is usually assessed by the decayed–missing–filled (DMF) index. The index represents the count of the number of either decayed (D), missing due to decay (M), or filled (F) teeth (T) or surfaces (S) and expresses the total caries and treatment experience of the individual at the time of observation. The index thus takes into account present decay (D), extracted teeth (M), and treated (F) lesions. Whether the teeth have been treated by extraction or filling is of little concern, as the basic idea behind the DMF system is the principle of adding past and present signs of caries. The chapter by Burt et al. (2008) in a textbook on dental caries offers useful background material for the discussion of the epidemiology of caries.

The DMF index could also serve as a basis for a discussion of the problem of analysing dental data, where non-independent measurements are often made, as a single individual may contribute up to 32 tooth-specific and 148 site-specific recordings of caries for the DMFT and DMFS indices, respectively. Furthermore, the distribution of caries among different teeth and surfaces has been shown to follow a certain pattern, with different surfaces showing different levels of susceptibility for developing the disease (Batchelor and Sheiham 2004). A useful point for a discussion of the many requirements for indices measuring dental diseases is the fact that the recorded disease occurrence comprises treatment components, designated M and F in the index, and that both are strongly dependent on the available dental services and the prevailing treatment philosophy. This is a good opportunity to review the basic definitions of measures of disease frequency and highlight that DMF is not just a measure of caries. This issue also raises conceptual questions in relation to the appropriateness of grouping together past and present disease and provides interesting comparisons with other medical conditions for which treated individuals are not necessarily grouped together with those currently experiencing the condition. Useful information can be obtained by looking at and analysing the DMF components separately, as D is indicative of current disease, M represents past disease treated surgically, and F indicates a conservative treatment approach to past disease. Students may also be asked to provide individual assessments of clinical slides or carry out clinical examinations of each other to illustrate problems such as definitions of lesions, observer variation, and the concept of validity and reliability. Finally, the analytical opportunities that the mouth offers,

such as split mouth trials (because caries occurs symmetrically on both sides of the jaw) or other types of paired designs, should also be mentioned (Vaeth and Poulsen 1998).

The epidemiology of dental caries is also important for demonstrating the links of available information with our understanding of the etiology and control of the disease (Murray 2003). The *association* between higher levels of fluoride in drinking water supplies and a lower caries experience was initially based on an ecological study. This observation was coincidental to studies investigating the 'Colorado Brown Stain' of teeth, or enamel mottling, now known as dental fluorosis. In the course of identifying fluoride as the element common to the water supplies that were associated with mottled enamel, it was noticed that there was less caries in communities with mottling (Dean et al. 1942). These results were used to determine 'optimal fluoride concentrations' of drinking waters (Hodge 1950) and were subsequently used as a basis for extensive artificial water fluoridation in many communities in the 1950s and 1960s.

A systematic review has shown that water fluoridation is indeed associated with a reduction in caries incidence and with an increased occurrence of dental fluorosis (McDonagh et al. 2000). A recent review indicated that the caries reductions are more modest in the last twenty years but still substantial (Rugg-Gunn and Do 2012). These findings offer a good basis for a discussion of the public health aspect of epidemiology, as well as equity issues and the health economic aspects of population preventive programmes.

A second large-scale ecological study on dental caries described the effect of wartime food rationing on the decline of dental caries (Takahashi 1961). Overall, there is historical evidence for a strong positive association between the consumption of refined carbohydrates and the occurrence of dental caries (Moynihan 2003). The occurrence of caries doubled between the eleventh and seventeenth centuries and increased dramatically during the late nineteenth century, when duties on sugar were removed. Aside from declines during the First and Second World Wars (Sheiham 1984), dental caries continued to increase until the middle of the twentieth century. The increases and decreases in dental caries that coincided with changes in sucrose consumption were interpreted to show a positive dose-response relationship and were given a causal interpretation. However, since the second half of the twentieth century, a strong trend for a caries decline has been noted (Marthaler 2004), and ecological studies now point to the absence of association (Woodward and Walker 1994) or even negative associations between sugar consumption and caries levels (Downer et al. 2008). These studies present a good opportunity to discuss the strengths and limitations of ecological studies, show their inappropriateness for the establishment of causal relationships, and help students understand the

concept of ecological fallacy. In addition, it is important to pay attention to the issue of causality (Scheutz and Poulsen 1999) and the application of the Bradford Hill guidelines for causal inference (Hill 1965); this can be done for caries but it is also worth attempting it for other important oral health conditions such as periodontal disease and oral cancer.

Data on caries from longitudinal studies can be used to highlight the important properties of these studies through demonstrating the progression of disease across childhood and adulthood, highlighting its cumulative nature, and familiarizing students with the concept of tracking of dental caries with clear and stable trajectories of the disease across the life course (Broadbent et al. 2008). The dramatic decline in dental caries in most industrialized countries also provides a good basis for introducing students to time series survey data and discussing the relative importance of fluoride in different forms, dietary changes, shifts in diagnostic criteria, and changing treatment paradigms. As it is the case for general health conditions, social factors, such as unemployment, poverty, and isolation have a considerable impact on oral diseases. This can be used to highlight the importance of social determinants of health (Marmot and Wilkinson 2006) and oral health (Marmot and Bell 2011) and present the basic concepts of social epidemiology. Because of the fact that many diseases, both oral and general, share common risk factors, this is also an opportunity to address the common risk factor approach (Watt and Sheiham 2012) and demonstrate how epidemiological information can be used to shape health promotion strategies and action (Daly et al. 2013).

Periodontal disease

The quantification of periodontal disease poses problems that are similar to those encountered when attempting to quantify dental caries: practically all persons present with at least mild signs of the disease. Consequently, comparison of fractions of the population affected by periodontal disease is of little value. Many attempts to establish periodontal indices have been made but the defining disease criteria remain disputed and periodontal disease diagnosis remains uncertain. Papapanou and Lindhe (2008) and Baelum and Lopez (2013) provide a discussion of the concept of these indices and of how they may have influenced our interpretation of epidemiological periodontal studies and formed dental health-care policies.

The widely held view of periodontal disease as a disease that starts as gingivitis and inevitably progresses to destructive periodontal disease if no intervention occurs has been challenged (Socransky et al. 1984; Papapanou and Lindhe 2008). The main criticism of the old model is the claim that epidemiological information and methods have been misinterpreted (Baelum and

Lopez 2013). Adult forms of destructive periodontal disease may be characterized by relatively short periods of exacerbation (bursts) separated by periods of remission lasting from a few days to a few years. This burst model necessitates a discussion of the precision of periodontal measurements in order to distinguish measurement error from real change, and this may be used as the basis for a discussion on the diagnostic validity and precision of continuous measurements of tissue destruction. Periodontal data may also be used to highlight regression-to-the-mean phenomena, as non-intervention studies may include repeated measurements of periodontal pockets (or other periodontal outcome measures) at very short time intervals, while intervention studies contain measurements before and after treatment. Due to measurement error and variation, initially higher estimates tend to be lower at a second measurement occasion and vice versa; therefore, using a correlation or regression technique to test the association between change in measurements and initial values may be problematic (Tu and Gilthorpe 2012). This presents a good opportunity to discuss such measurement and analytical issues and illustrate the superiority of the Bland–Altman method for comparisons (Bland and Altman 1986).

A characteristic feature of periodontal disease is the distribution of lesions in affected people. This shows that the periodontal tissues adjacent to the 128 surfaces of teeth at risk are not at equal risk and that the disease does not affect sites at random. This has implications for periodontal disease estimates, since it is common in epidemiological studies to measure disease only at selected sites on a few teeth, a so-called partial mouth recording, which may bias disease estimates. Finally, clinical epidemiological studies of periodontal disease can be used to spark discussions on how misinterpretations and incorrect statistical analysis can lead to major health policy and treatment decisions, hence pointing out to the necessary balance between statistical and clinical significance (Imrey 1986; DeRouen et al. 1995; Addy and Newcombe 2005).

The manifestations of periodontal disease are influenced by different risk factors, some of which, such as stress and smoking, are strongly associated with other general health conditions, such as cardiovascular disease. This discussion thus provides a good opportunity to highlight the importance of the common risk factor approach and the integration of oral health promotion into general health promotion strategies (Watt and Petersen 2012). On a separate point, the association that may be observed between periodontal disease and a number of general health conditions such as cardiovascular disease may facilitate a useful discussion about causal inference, confounding, and ways of controlling confounding (Lockhart et al. 2012).

Screening for oral diseases

The concept of screening is important to discuss in the dental context, since the typically semi-annual routine recall visit to the dentist is a well-known example of a screening examination. Obviously, this provides the setting for a discussion of the scientific basis for screening, the WHO criteria for screening, and the public health aspects of such screening programmes (Wilson and Jungner 1968), particularly as there is insufficient evidence about the value of such routine recall programmes (Beirne et al. 2005). Furthermore, the impact of the less-than-perfect reproducibility and accuracy of the disease detection methods should be discussed and particular attention directed to consequences of the poor sensitivity and specificity of caries and periodontal disease detection methods. Students should be trained in entering hypothetical data into a diagnostic decision matrix in order to calculate and understand the importance of positive and negative predictive values of a diagnostic test, and especially how these values depend on disease prevalence (Hausen 1997). Hypothetical or 'real' caries data may be used to construct a curve showing the positive predictive values at different possible prevalence values. The consequences concerning treatment in general and dental caries treatment in particular should be discussed, especially in view of the decline in caries.

Epidemiological information as a tool for planning interventions

Epidemiological methods are used for identifying causes of diseases and establishing priorities for health policy and action. In that respect, the collection and interpretation of epidemiological information on needs assessment becomes crucial for the provision and organization of dental health services. In most settings, oral health needs have been exclusively determined through normative (clinical) need, that is, through clinical assessment by professionals. However, the sole use of clinical need has considerable limitations and cannot appropriately reflect the true levels of need (Sheiham and Tsakos 2007). Furthermore, it is widely recognized that health-care needs may be defined in different ways that do not just reflect the clinical reality. After all, the dental-care needs of an individual are not only defined through clinical status, but also via other factors such as the oral health perceptions and wants of the specific person. Therefore, the socio-dental approach to needs assessment was proposed in order to overcome the limitations of the sole use of normative need assessments, by also gradually incorporating subjective perceptions of the related oral impacts as well as behavioural propensity factors, while considering evidence-based interventions (Gherunpong et al. 2006; Sheiham and Tsakos 2007).

Students can carry out small surveys on people to assess their oral health needs. The gap between normative and perceived needs can be demonstrated by interviewing respondents about their needs. Furthermore, students have the opportunity to come to terms with the non-clinical aspects of oral health and become familiar with subjective measures of oral health status, including oral health related quality of life measures (Locker and Allen 2007). Here, medical sociology and health economics can be integrated by discussing the concepts of health and illness and putting the professional and lay views in a broader context. In addition, the discussion about oral health related quality of life measures could be linked with a methodological discussion about the measurement and interpretation of scores, thereby advancing the already acquired statistical knowledge about presentation and distribution of disease (Tsakos et al. 2012).

Epidemiology should be used to assist in making decisions on priorities and strategies. The concepts of high-risk and population strategies (Rose 1992; Rose et al. 2008) are used to illustrate how epidemiology and policy are interrelated. This is done by applying Rose's concepts to the current policy debate in dentistry, namely, whether the main emphasis should be on detecting and treating those at higher risk of disease rather than directing resources at the whole community to lower disease levels in general (Sheiham and Joffe 1991; Baelum 2011). The risk approach can attempt to identify either population subgroups (known as the targeted-population approach) or individuals (the high-risk approach).

The advantages and disadvantages of the two approaches can be outlined (Daly et al. 2013). The implication of the high-risk strategy is that one should reach those most in need of intervention. This approach invites the question 'How can we identify them?' Based on the fact that most new caries lesions occur in children classified as low risk at baseline, Batchelor and Sheiham (2006) questioned the rationale for the application of the high-risk approach. Another question is the ability of any method of screening to separate high and low risk groups, in terms of sensitivity, specificity, and predictive values. Just as the high-risk strategy requires a scientific basis both in technical matters (methods of identifying those at high risk) and evaluation (validity, effectiveness), the same is true of the whole population strategy.

Rose (1992) has eloquently discussed the scientific basis for the whole population strategy. He draws the distinction between two kinds of etiological questions. The first seeks the causes of cases: 'Why do some people get caries at this time?' The second seeks the causes of the incidence: 'Why do some populations have much caries while in others it is uncommon?' The whole population strategy attempts to control the determinants of the incidence and is therefore a more radical strategy. It depends upon epidemiological, sociological, and

other kinds of research in order to first identify important determinants of the diseases in question and then act to change them in the appropriate direction. This, in turn, is central to the required action needed to tackle oral health inequalities.

Teaching methods and format

The key concepts and items to be taught on an undergraduate level can be covered during a short-term course (Tables 23.1 and 23.2), while at postgraduate level this can vary according to the specific field of study. For example, for an MSc in restorative dentistry, there will be a generally limited allocation of sessions for both statistics and epidemiology, while in a respective course in dental public health the teaching of epidemiology should cover at least one whole module and there may be another two modules for statistics (basic and advanced). The undergraduate course can take place at a specific time in the curriculum or be integrated into different themes. All teaching material is distributed to the students before the course starts. The material comprises relevant scientific articles, questions focusing on important aspects of the articles, small independent exercises based on real datasets, and one larger exercise to be completed on an individual basis after completion of the course. The background material consists of books covering the basics in biostatistics and epidemiology, and copies of a few relevant publications. To facilitate the learning process, the core elements of the biostatistics and epidemiology should also be available on an intranet page made available to the students. Core lectures should be given during the course. The exercises should be completed as group work and, after the completion of each exercise, a group should be selected to present and discuss the answers of the exercise in a plenary session.

Assessing students' achievements

All teachers monitor the students' achievements throughout the course by their ability to complete the exercises, the solutions produced by each group, and their contribution in the discussions. In the final evaluation we test the students as critical consumers of dental literature by asking each student to produce a critical review of a scientific report on a clinically relevant problem. This review is subsequently evaluated with respect to the degree of fulfilment of the course goals. The described format and teaching methods facilitate teaching dental students basic statistics and epidemiological methods and at the same time aim to transform dental students into critical consumers of scientific dental research.

References

Addy, M. and Newcombe, R. G. (2005) Statistical versus clinical significance in periodontal research and practice. *Periodontology 2000*, **39**: 132–44.

Baelum, V. (2011) Dentistry and population approaches for preventing dental diseases. *Journal of Dentistry*, **39 Suppl.** 2: S9–19.

Baelum, V. and Lopez, R. (2013) Periodontal disease epidemiology – learned and unlearned? *Periodontology 2000*, **61**: 37–58.

Batchelor, P. A. and Sheiham, A. (2004) Grouping of tooth surfaces by susceptibility to caries: a study in 5–16 year-old children. *BMC Oral Health*, **4**: 2.

Batchelor, P. A. and Sheiham, A. (2006) The distribution of burden of dental caries in schoolchildren: a critique of the high-risk caries prevention strategy for populations. *BMC Oral Health*, **6**: 3.

Beirne, P., Clarkson, J. E., and Worthington, H. V. (2005) Recall intervals for oral health in primary care patients. *Cochrane Database Systematic Review*, **18**: CD 004346.

Biggs, J. (2003) *Teaching for Quality Learning at University* (2nd edn). Berkshire: Open University Press.

Bland, J. M. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, **327**: 307–10.

Bonita, R., Beaglehole, R., and Kjellström, T. (2006) *Basic Epidemiology* (2nd edn). Geneva: World Health Organization.

Broadbent, J. M., Thomson, W. M., and Poulton, R. (2008) Trajectory patterns of dental caries experience in the permanent dentition to the fourth decade of life. *Journal of Dental Research*, **87**: 69–72.

Burt, B. A., Baelum, V., and Fejerskov, O. (2008) 'Epidemiology of dental caries', in **O. Fejerskov** and **E. A. M. Kidd**, eds, *Dental Caries: The Disease and its Clinical Management* (2nd edn). Oxford: Blackwell Munksgaard, pp. 127–50.

Coggon, D., Rose, G., and Barker, D. J. P. (2003) *Epidemiology for the Uninitiated*. (5th edn). London: BMJ Books.

Daly, B., Batchelor, P., Treasure, E. T., and Watt, R. G. (2013) *Essential Dental Public Health* (2nd edn). Oxford: Oxford University Press.

Dean, H. T., Arnold F. A. Jr., and Elvove, E. (1942) Domestic water and dental caries. V. Additional studies of the relation of fluoride domestic waters to dental caries experience in 4,425 while children aged 12–14 years of 13 cities in 4 states. *Public Health Reports*, **57**: 1155–79.

DeRouen, T. A., Hujoel, P. P., and Mancl, L. A. (1995) Statistical issues in periodontal research. *Journal of Dental Research*, **74**: 1731–7.

Downer, M. C., Drugan, C. S., and Blinkhorn, A. S. (2008) Correlates of dental caries in 12-year-old children in Europe: a cross-sectional analysis. *Community Dental Health*, **25**: 70–8.

Gherunpong, S., Sheiham, A., and Tsakos, G. (2006) A sociodental approach to assessing children's oral health needs: integrating an oral health-related quality of life (OHRQoL) measure into oral health service planning. *Bulletin of the World Health Organization*, **84**: 36–42.

Greenberg, R. S., Daniels, S. R., Flanders, W. D., Eley, J. W., and Boring, J. R. (2004) *Medical Epidemiology* (4th edn). New York: Lange Medical Books/McGraw-Hill.

Greenhalgh, T. (1998) *How to Read a Paper: The Basics of Evidence Based Medicine*. London: BMJ Publishing Group.

Hausen, H. (1997) Caries prediction—state of the art. *Community Dentistry and Oral Epidemiology*, **25**: 87–96.

Hennekens, C. H., Buying, J. E., and Mayrent, S. L. (1987) *Epidemiology in Medicine*. Boston, MA: Lippincott Williams & Wilkins.

Hill, B. (1965) The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine*, **55**: 295–300.

Hodge, H. C. (1950) The concentration of fluorides in drinking water to give the point of minimum caries with maximum safety. *Journal of the American Dental Association*, **40**: 436–9.

Imrey, P. B. (1986) Considerations in the statistical analyses of clinical trial in periodontics. *Journal Clinical Periodontology*, **13**: 517–28.

Kirkwood, B. R. and Sterne, J. A. C. (2003) *Essential Medical Statistics* (2nd edn). Oxford: Blackwell Science.

Locker, D. and Allen, F. (2007) What do measures of 'oral health-related quality of life' measure? *Community Dentistry and Oral Epidemiology*, **35**: 401–11.

Lockhart, P. B. et al. (2012) Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association? A scientific statement from the American Heart Association. *Circulation*, **125**: 2520–4.

Marmot, M. and Bell, R. (2011) Social determinants and dental health. *Advances in Dental Research*, **23**: 201–6.

Marmot, M. and Wilkinson, R. (2006) *Social Determinants of Health* (2nd edn). Oxford: Oxford University Press.

Marthaler, T. M. (2004) Changes in dental caries 1953–2003. *Caries Research*, **38**: 173–81.

McDonagh, M. S., Whiting, P. F., Wilson, P. M., Sutton, A. J., Chestnutt, I., Cooper, J., Misso, K., Bradley, M., Treasure, E., and Kleijnen, J. (2000) Systematic review of water fluoridation. *British Medical Journal*, **321**: 855–9.

Moynihan, P. (2003) 'Diet and dental caries', in J. J. Murray, J. H. Nunn, and J. G. Steele, eds., *The Prevention of Oral Disease* (4th edn). Oxford: Oxford University Press, pp. 7–34.

Muir Gray, J. A. (1997) *Evidence-based Healthcare*. Edinburgh: Churchill Livingstone.

Murray, J. J. (2003) 'Fluorides and dental caries', in J. J. Murray, J. H. Nunn, and J. G. Steele, eds, *The Prevention of Oral Disease* (4th edn). Oxford: Oxford University Press, pp. 35–60.

Papapanou, P. N. and Lindhe, J. (2008) 'Epidemiology of periodontal disease', in J. Lindhe, N. P. Lang, and T. Karring, eds, *Clinical Periodontology and Implant Dentistry* (5th edn). Oxford: Blackwell Munksgaard, pp. 129–80.

Rose, G. (1992) *The Strategy of Preventive Medicine*. Oxford: Oxford University Press.

Rose, G., Khaw, K-T., and Marmot, M. (2008) *Rose's Strategy of Preventive Medicine*. Oxford: Oxford University Press.

Rothman, K. (2002) *Epidemiology: An Introduction*. Oxford: Oxford University Press.

Rugg-Gunn, A. J. and Do, L. (2012) Effectiveness of water fluoridation in caries prevention. *Community Dentistry and Oral Epidemiology*, **40** Suppl. 2: 55–64.

Scheutz, F. and Poulsen, S. (1999) Determining causation in epidemiology. *Community Dentistry and Oral Epidemiology*, **27**: 161–70.

Sheiham, A. (1984) Changing trends in dental caries. *International Journal of Epidemiology*, **13**: 142–7.

Sheiham, A. and Joffe, M. (1991) 'Public dental health strategies for identifying and controlling dental caries in high and low risk populations', in N. W. Johnson, ed, *Markers of High and Low Risk Groups and Individuals for Dental Caries*. Cambridge: Cambridge University Press, pp. 445–81.

Sheiham, A. and Tsakos, G. (2007) 'Oral health needs assessment', in C. M. Pine and R. Harris, eds, *Community Oral Health*. New Malden: Quintessence, pp. 59–79.

Sheiham, A. and Watt, R. G. (2003) 'Oral health promotion and policy', in J. J. Murray, J. H. Nunn, and J. G. Steele, eds, *The Prevention of Oral Disease* (4th edn). Oxford: Oxford University Press, pp. 241–57.

Socransky, S. S., Haffajee, A. D., Goodson, J. M., and Lindhe, J. (1984) New concepts of destructive periodontal disease. *Journal of Clinical Periodontology*, **11**: 21–32.

Takahashi, K. (1961) Statistical study on caries incidence in the first molar in relation to the amount of sugar consumption. *Bulletin of Tokyo Dental College*, **1**: 58–70.

Tsakos, G., Allen, P. F., Steele, J. G., and Locker, D. (2012) Interpreting oral health-related quality of life data. *Community Dentistry and Oral Epidemiology*, **40**: 193–200.

Tu, Y. K. and Gilthorpe, M. S. (2012) Key statistical and analytical issues for evaluating treatment effects in periodontal research. *Periodontology 2000*, **59**: 75–88.

Væth, M. and Poulsen, S. (1998) Comments on a commentary: statistical evaluation of split mouth caries trials. *Community Dentistry and Oral Epidemiology*, **26**: 80–3.

Watt, R. G. and Petersen, P. E. (2012) Periodontal health through public health – the case for oral health promotion. *Periodontology 2000*, **60**: 147–55.

Watt, R. G. and Sheiham, A. (2012) Integrating the common risk factor approach into a social determinants framework. *Community Dentistry and Oral Epidemiology*, **40**: 289–96.

Wilson, J. M. G. and Jungner, G. (1968) *Principles and Practice of Screening for Disease*. Public Health Paper No. 34. Geneva: WHO.

Woodward, M. and Walker, A. R. (1994) Sugar consumption and dental caries: evidence from 90 countries. *British Dental Journal*, **176**: 297–302.

Chapter 24

Clinical epidemiology

John A. Baron, Henrik Toft Sørensen, and Harold C. Sox Jr

Introduction to clinical epidemiology

The term 'clinical epidemiology' has been interpreted in various ways but the core of the discipline is generally understood to be the application of epidemiological and biostatistical techniques to clinical problems. In contrast to etiologic epidemiology, which focuses on the determinants of disease on a general population level, clinical epidemiology addresses the outcomes of disease and illness in clinical populations. The goal is a practical one: to aid clinicians in their work diagnosing, advising, and treating patients. 'Evidence-based medicine' is a related concept, the integration of current best evidence and clinical expertise in the care of individual patients. Thus, evidence-based medicine is essentially the application of clinical epidemiology to the daily work of caring for individual patients.

Teaching objectives

Naturally, the topics covered in a clinical epidemiology course will depend on the level and context in which it is taught. Some students will be medical students or clinicians seeking to better understand the clinical literature and improve their clinical practices. Many of these students will have only a limited quantitative background. For them, a clinical epidemiology course should foster awareness of the epidemiological and statistical concepts that bear on clinical issues, improve understanding of the clinical medical literature, and provide cognitive skills that will help in the care of patients. More advanced students will be budding investigators who want to develop capabilities for clinical research; a course for this group would naturally assume a broader understanding of epidemiology and biostatistics in general. This chapter will focus on the content of the basic course, indicating additional advanced topics, as appropriate.

A basic but fairly comprehensive course should provide the students with an understanding of the quantitative approaches that characterize clinical epidemiology, and include most of the following topics:

- basic aspects of medical measurement—sources of variation and the consequences of that variability;
- measures of disease occurrence and trait/outcome associations;
- an introduction to clinical scales and prediction rules;
- the fundamentals of research design, including the strengths and weaknesses of various observational designs, clinical trials, and the concepts of bias and confounding;
- the application of these designs to clinical problems such as estimation of prognosis and evaluation of treatment efficacy;
- use of probability in the interpretation of clinical tests;
- expected value decision analysis; and
- quantitative aspects of screening and prevention.

More advanced topics within these areas are summarized in the corresponding discussions below. For advanced students, an important goal should be reinforcing rigorous statistical and epidemiological thinking regarding the topics covered.

Teaching content

Quantitative concepts

It is essentially impossible to teach clinical epidemiology without using concepts of probability and statistics, which are needed to deal with a central reality of clinical medicine: the pervasiveness of uncertainty. This is obvious in risk factor research: even a very strong risk factor (e.g. smoking for lung cancer) does not predict disease occurrence well (Pepe et al. 2004). Similarly, there is uncertainty inherent in clinical associations and decision making: few treatments are effective in all patients, and even some patients with a poor prognosis will do well. In a basic course the relevant quantitative concepts can be reinforced at natural junctures, and elements of a few more advanced biostatistical ideas can be introduced (e.g. survival analysis).

The use of probability is central to clinical epidemiology. A basic concept is risk, the probability of a future event. Risks and rates are distinct but related concepts related to the incidence of a disease or outcome. Estimation of either requires a well-defined numerator that corresponds appropriately to its denominator, as well as suitable consideration of censoring and (for risks) a specified

time horizon. Technical discussions regarding their estimation and interrelationships (Morgenstern et al. 1980) are often left to more advanced courses. Probabilities are also used to describe the prevalence of a condition or disease, as this is inherently a cross-sectional issue. Nonetheless, these probabilities are sometimes (improperly) referred to as 'risks'.

The appropriate use and interpretation of ratio and difference measures of association is a fundamental topic. Traditional (etiologic) epidemiology generally uses relative measures of association, while clinical epidemiology, with its focus on clinical impact, also requires absolute risks, or the difference of risks in two exposure or treatment groups. This class of statistics is essential for discussions of prognosis and for assessment of the benefits and harms of medical interventions. For preventive or treatment exposures, the number needed to be treated is one way to summarize effectiveness of the intervention. This is the reciprocal of the risk difference and expresses the number of subjects that need to be treated on average to avoid one end-point event. The number needed to harm is similarly defined for adverse exposures. Details of the distinctions between the various types of relative risks (e.g. rate ratios, risk ratios, odds ratios, prevalence ratios, etc.) can be relegated to an advanced course, as can treatment of life expectancy and years of life lost (or gained; Wright and Weinstein 1998).

Medical measurement

Components of sickness; concepts of biostatistics

The distinction between disease and illness (Taylor 1979) illustrates many fundamental measurement issues. The disease (the physiologic or psychological disturbance underlying the sickness) often has associated numeric measurements (e.g. the area of a stenotic valve). In contrast, the illness (the patient's experience of the disease) is often measured with ordinal or nominal assessments. These are difficult measurements: scales will be needed, and, typically, there is no gold standard. The relationship between disease and the associated illnesses is imperfect and, with molecular methods of disease detection, the distinctions are likely to become even wider (Black and Welch 1993). Discussion of labelling (Alderman and Lamport 1990) reinforces the differences between disease and illness.

Validity and precision; measures of agreement and disagreement

All medical measurements have the potential for variation, some of which is due to variability in the underlying biology (within and between-person variation) and some due to the measurement process (measurement variation). Interpreted tests such as X-rays and anatomic pathology may be particularly

prone to the latter. Some striking examples of disagreement in interpreted tests can motivate consideration of the sources of variation in clinical data. There are some (now dated) papers that have summarized data regarding the repeatability of various clinical measures (Elmore and Feinstein 1992). Stage migration (the 'Will Rogers phenomenon'; Feinstein et al. 1985; Black and Welch 1993) is a consequence of variation in staging measurements.

Validity and precision are terms often used to describe the variation (or lack of variation) in measurements. Validity (sometimes referred to as 'accuracy') refers to the tendency of data to measure what is intended; statistically, this is lack of bias. Precision is simply the repeatability of the measurement (i.e. small variability). The cardinal characteristic of 'hard' data is precision (i.e. repeatability). Regression toward the mean (Barnett et al. 2005) is a common, often unrecognized, consequence of measurement variability. In clinical settings, it can lead to false impressions of treatment efficacy and a distorted understanding of the natural history of disease. Discussion of regression to the mean naturally leads to a consideration of the effects of subject selection and the need for controls, topics that can be returned to during discussions of research designs.

Understanding clinical measurement involves issues of agreement and disagreement between measurements (Bland and Altman 1995). These topics can force a review of some basic quantitative concepts and can also be the occasion for the introduction of new statistics (e.g. kappa statistics). Although the rationale for this measure is fairly easy to explain, its interpretation is not clear when there are more than two observers, if the measurement is not dichotomous, or if the distribution of measurements is strongly skewed (Maclure and Willett 1987). The continuous counterpart, the intraclass correlation, is usually not included in first courses. The term can refer to several quantities but probably can best be used as the ratio of the between-person variance to the total variance of a set of measurements.

After the theme of variability is developed, issues of normal ranges of clinical measurements can be discussed. The term 'normal' carries several implications that can usefully be explored: Gaussian distribution, usual value, and desirable value. The idea of a normal test value can be picked up later, during a discussion of clinical testing.

Clinical scales

In medicine, clinical scales are summary measures of a state of disease or health, such as those describing functional status, quality of life, or symptoms. These typically combine several different measurements into a single index. Cancer staging is perhaps the best known example. Other commonly used scales include the Apgar score and the ASA physical status classification system.

Utilities (Sox et al. 2013) are a sort of scale that has particular relevance for decision analysis.

Understanding clinical scales (McDowell 2006) involves many of the measurement issues mentioned above and brings in concepts such as face validity, criterion validity, and construct validity. Often clinical scales will address the illness dimension of sickness; the validation of such scales will require statistics to describe scale coverage, reliability, validity, responsiveness/sensitivity, and calibration (McDowell 2006). These topics, as well a discussion of the construction of scales and utilities (Streiner and Norman 2008), will probably best be left to an advanced course.

Research design

Research designs are considered in detail elsewhere in this volume (see chapter 3). This discussion will consider only those aspects that are particularly relevant for teaching clinical epidemiology. Although all of the classic epidemiologic research designs may be used to address clinical questions, their use may be quite different than in etiologic epidemiology. Cross-sectional studies are not well suited for etiologic investigation but are the design of choice for clinical questions such as the comparison of clinical measurements or studies of diagnostic tests. Case-control studies are less commonly used in clinical epidemiology than in etiologic studies.

There are several design features in which etiologic epidemiology and clinical epidemiology tend to differ. While broad population-based studies are particularly valuable in etiologic research, studies of diagnosis, treatment, and prognosis may easily be and validly conducted in clinical populations. Thus medical databases (see chapter 26) are particularly useful in clinical studies. In addition to investigating disease incidence and mortality, the focus of most etiologic research, clinical epidemiological studies may focus on disease recurrence, quality of life or symptom burden. Propensity scores are often used to control confounding in studies of treatments. In contrast to etiologic studies, confounding is irrelevant to many areas of clinical epidemiology such as risk prediction and diagnosis. What matters is the accuracy of the prediction, not whether there is a causal association with an individual exposure or trait.

There are some biases that are particularly important for clinical epidemiology. A classic example is confounding by indication (Walker 1996), which may occur in the observational assessment of the association between medical treatments and various outcomes. The treatments are generally given for a reason (the 'indication'), which has an association with adverse outcomes and thus becomes a confounding factor. The idea that use of anti-hypertensive agents is associated with stroke risk is an extreme example. Other biases may occur

because of the selection of patients into particular medical practices or selection associated with hospitalization.

Prognosis studies (Hemingway 2006) are an excellent entry into research design. The natural history of a disease is its untreated prognosis; the clinical history is its prognosis under medical care (Fletcher et al. 2012). A prognosis is essentially a set of probabilities of various outcomes over time, so the fundamental language of prognosis is naturally that of risk. Concepts and techniques of survival analysis will need to be introduced or reviewed:

- the concept of risk as a probability for a person, assessed by the average risk (measured in a group);
- the dependence of event probabilities on the length of follow-up, censoring, and competing risks;
- the limitations of describing survival only over one time interval (e.g. a five-year survival rate) and the advantages of Kaplan–Meier and cumulative incidence curves;
- the desirability of considering prognosis from a defined point in the course of the disease, preferably at the time of incidence (i.e. the formation of an inception cohort); and
- the need to compare the outcomes for patients having a certain disease with some familiar 'benchmark'—this leads naturally to the concept of controls and cohort studies.

The natural design for prognostic investigation is a cohort study. Unbiased follow-up will be needed for effective comparison of the exposure groups. In contrast to etiologic epidemiology, where the risk factors are often lifestyle factors, prognostic factors are dominated by characteristics of the disease (e.g. stage of tumour, extent of or location of atherosclerosis), its treatment, or its co-morbidities. Case-control studies can be considered as an alternative design for assessing the magnitude of the same exposure/outcome association, but these are far less useful in this context since absolute and attributable risks cannot be directly computed. The limitations of these observational designs need to be considered, including issues of chance, confounding, and study-related bias. The fact that exposures are interrelated in real life sets up the conditions for confounding.

Clinical trials (Machin and Fayers 2010) deserve special emphasis in a clinical epidemiology course because of their importance in clarifying clinical issues of treatment, screening, and prevention. Trials can be seen as a special type of cohort study, with random assignment of the exposure and (usually) a highly organized follow-up. Randomization has an important role as statistical protection against

confounding, even against potential confounding factors that are not measured or unknown. The powerful role of randomization cannot be duplicated with observational techniques such as matching or adjustment. In order to preserve the advantages of randomization, a clinical trial analysis needs to be based on the randomized treatment groups, including in each treatment group subjects that drop out or even cross over to other study treatments. This may seem counter-intuitive or even incorrect but can be explained as the only analysis based entirely on the groups created by randomization. The 'as-treated' analysis studies subjects according to the treatments actually used. This introduces observational elements and so is properly a secondary analysis. Even farther from the intention-to-treat approach is the per-protocol analysis, which includes only subjects who fulfil the most important elements of the protocol: those who take only the assigned treatment with a specified level of compliance, are not examined outside the stipulated time windows, etc. The value of a low drop-out rate follows directly; this makes the intention-to-treat analysis and the as-treated analysis very similar. To the extent that there is non-compliance, the intention-to-treat analysis will usually provide a conservative estimate of treatment effects (Sorensen et al. 2006). Blinding is a second procedure that provides important advantages for clinical trials, these being avoidance of bias in the assignment of treatments ('allocation concealment') and in the assessment of end points.

Clinical trials, particularly multicentred trials, are typically relatively formal investigations, with publicized pre-study hypotheses, very detailed protocol and operations manuals, an independent safety and data-monitoring committee, etc. These formalisms can seem bureaucratic and unnecessary unless their role in preserving the validity of the associated study is made clear. This effort will then serve to reinforce many of the basic principles of clinical trial research.

Special types of trials (Machin and Fayers 2010) can be discussed if there is time, or left to a more advanced course. Equivalence trials aim to test whether treatments are essentially equipotent, and often require large sample sizes. Cross-over trials allow each person to serve as a 'self-control' and so introduce some efficiencies. Prevention trials can be difficult because they often need to be continued for many years, and the study interventions may be widely available to the participants, leading to crossing-over of placebo subjects to active treatment. 'N of 1' trials are occasionally used to identify effective treatments for individual subjects. Cluster trials are useful for comparing health care delivery strategies because the unit of randomization is the clinical unit (e.g. a clinical practice).

Many aspects of clinical trials can be accommodated in observational research. Randomization is essentially unique to trials, although in some circumstances randomness in real life can be taken advantage of. Genes are randomly

assorted at meiosis, creating the opportunity for 'Mendelian randomization' (Lawlor et al. 2008). More commonly, other trial elements can be used to enhance observational studies, including careful choice of study population, blinded ascertainment of outcomes, and organized follow-up (stipulated endpoint measurements at predetermined times). These opportunities can be mentioned in basic courses and pursued in more detail in advanced courses.

Findings from clinical trials and observational research are often in broad agreement (Sorensen et al. 2006). However, recent examples of serious discordance include the effects of menopausal hormones on risk of cardiovascular disease, and the purported protective effect of beta-carotene intake on the risk of lung cancer. Examples such as this are very productive discussion points, since they force detailed considerations of the research designs and the manner in which they are implemented. In some cases, differences between clinical trial and observational research may be due to the tendency of the two disciplines to study different things. Many trials aim to investigate the 'efficacy' of the intervention: its effects under more or less ideal conditions, with careful choice of subjects, excellent compliance, etc. Observational studies tend to reflect the 'effectiveness' of the intervention: its effects in the real world, with use in a broader range of patients and variable compliance.

By tradition (though certainly not necessity), several topics of general importance are often presented in the context of clinical trials, including data dredging, generalizability (internal vs external validity), subgroup analyses, and meta-analyses. A discussion of these issues can serve to review important aspects of research design and interpretation for virtually all types of studies.

Prediction rules

Clinical prediction rules are structured statistical methods for estimating a probability that a condition is present or the risk of a future disease outcome. Widely used clinical prediction rules include the Geneva rule for estimating the probability that a pulmonary embolism is present, APACHE scores of critical illness, the Gail model of breast cancer risk, and the Framingham risk predictions for cardiovascular disease.

The basic designs that underlie prediction modelling are simple: cross-sectional studies for estimation of disease prevalence, and cohort studies for the estimation of disease risk. However, the derivation of the models can be extremely complex statistically (involving various types of regression modelling, recursive partitioning, etc.). Details will be inappropriate for a beginning course; even a more advanced, general clinical epidemiology course can only touch on that topic. A more accessible—and more important—issue for the

user of a proposed clinical decision rule is its validation (Wasson et al. 1985). The multivariate predictions can be derived on a training dataset and tested on a separate dataset, or evaluated using procedures such as bootstrapping or cross-validation. These matters can also be quite technical (and so unsuitable for a first course) but the underlying principles of validation should be emphasized. A prediction rule constructed in one population will need to be validated (and possibly recalibrated) for use in other populations.

There are two different aspects of the accuracy of prediction rules. Calibration (sometimes referred to as 'reliability') refers to the accuracy of the prediction, that is, whether the rule predicts the same proportion of positives in subgroups of the training set and the test set. Discrimination, on the other hand, refers to the ability to distinguish individuals who will have the outcome from those who will not. In theory, a prediction rule may show excellent discrimination but poor calibration, and vice versa. For diagnosis, discrimination is the important parameter. However in other settings calibration may be key: reproducible prediction of disease risks is needed for many clinical decisions. Advanced courses can consider the various measures of discrimination and calibration. The area under the receiver operating characteristic (ROC) curve, a measure of discrimination, is not directly helpful in understanding calibration.

Diagnosis and testing

Overview

The diagnostic process should occupy a significant proportion of most clinical epidemiology courses, since here the consequences of uncertainty are well known and the advantages of quantitative reasoning are quite obvious (Sox et al. 2013). A brief account of the basics of differential diagnosis is good starting point, to show where qualitative reasoning ends and probabilistic reasoning can begin. The experienced diagnostician uses several qualitative and diagnostic strategies, including pattern recognition and branching algorithms, to narrow diagnostic possibilities and focus on individual diseases. Formal testing theory can inform the pivotal decision of ruling out or ruling in a diagnosis. This approach involves the use of Bayes' rule to combine test operating characteristics (such as sensitivity and specificity or likelihood ratios) and pre-test probability to estimate a post-test probability. New information (test results) moves pre-test probabilities up or down in a way that accurately reflects the contribution of the new data. As described below, the course should provide beginning students with the ability to understand and distinguish these different types of probabilities.

Estimating pre-test probabilities

Pre-test probabilities can be assessed objectively by observing the frequency of a disease in patients with a clinical findings or using clinical prediction rules. However, in practice, probability estimation is often subjective and based largely on relatively informal impressions of prior experience (Sox et al. 2013). The heuristics (mental shortcuts) that people use to subjectively estimate probabilities are subject to error. The three principal heuristics are representativeness (the more an event looks like a typical case, the higher the assigned probability), anchoring and adjustment (failure to adjust probabilities sufficiently when incorporating new information), and availability (judging the probability of an event by how easily it comes to mind). For these reasons, quantitative estimation of probabilities is preferred when the data exist to support it.

Test operating characteristics

A beginning course should focus on developing a clear understanding of sensitivity and specificity, index test, gold standard test, and the cut-point between 'positive' and 'negative' tests. A more advanced course can discuss the variation in sensitivity and specificity, how to measure them, and how to deal with uninterpretable test results. The most important bias in estimation of test operating characteristics is test-referral, which occurs when patients with a positive index test result are referred more often for the diagnostic reference standard test than those with a negative test result.

The most basic treatment of testing theory assumes a dichotomous test result. However, most test measurements are fundamentally multichotomous or continuous. Consequently, in many clinical circumstances, 'normal' versus 'abnormal' will not capture all the information inherent in the test result. Likelihood ratios are convenient summaries of test operating characteristics for these test measurements: the probability of a given test result in persons with the target disease divided by the probability in those without the disease (Sox et al. 2013). An ROC curve displays the sensitivity and (1-specificity) of all possible test results; the area under the curve expresses the inherent ability of the measurement to discriminate diseased from nondiseased patients.

Computing and interpreting post-test information: Bayes' theorem

Bayes' theorem yields the post-test probability of a target condition given a particular test result and the sensitivity and specificity of the test measurement. The use of Bayes' theorem rests on the assumption that the sensitivity and specificity are inherent characteristics of the test, independent of the pre-test probability. A key point is that, with fixed sensitivity and specificity, the post-test

probabilities vary with the prior probability and are not inherent features of the test measurement. This point bears emphasis, with the use of examples in which the disease prevalence is varied and the post-test probabilities change with the prior probabilities. This variation in the posterior probabilities with the prior probability may be counter-intuitive to some clinicians, since it means that the meaning of a positive (or negative) test result (i.e. post-test probabilities) will vary, depending on the clinical circumstances (i.e. the prior probability). There are several distinct clinical settings for the interpretation of a test (e.g. screening, ruling out disease, and ruling in disease). These tend to have different ranges of pre-test probabilities, and different consequences of false-positive and false-negative errors. The final message is that, to rule out disease, a negative result of a very sensitive test will be needed, while to rule in disease, a positive result on a relatively specific test will be required.

For teaching purposes, the discussion of Bayes' theorem can usefully feature its odds ratio form:

$$\text{post-test odds} = \text{pre-test odds} \times \text{likelihood ratio}.$$

In addition to making calculation of the post-test odds relatively easy, this formulation focuses attention on the structure of the diagnostic process, neatly separating pre-test estimates, test operating characteristics, and post-test estimates. It also serves to emphasize the importance of the likelihood ratio in the diagnostic process.

After applying a (dichotomous) test and the diagnostic reference standard to a patient population, a 2×2 table may be created, with test results on (say) the rows, and disease status on (say) the columns. Sensitivity and specificity can then be computed from data within separate columns. Data within the rows can be used to compute the probability of disease among those who test positive (the positive predictive value of the test) and the probably of absence of disease among those who test negative (negative predictive value). For this specific patient population, these predictive values equal the post-test probabilities after a positive or negative test, as computed with Bayes' theorem. These predictive values are empirical numbers, derived from using the test in a particular patient population. These features of the 2×2 data are both a strength (no need to assume constancy of test performance) and a weakness (the predictive values apply only to the population studied).

Topics for more advanced courses include mathematical aspects of Bayesian testing (Baron 1994; Sox et al. 2013) and the statistics of ROC curves (e.g. estimation of the area under the curve, the testing of differences in the areas for two different test measurements, and the identification of optimal cut-points between 'positive' and 'negative' test results).

Screening and prevention

At the onset of a discussion of screening (Morrison 1992; Black and Welch 1993), the instructor should draw the distinction between screening and diagnostic testing: the objective of screening is to detect disease before symptoms occur. Therefore, screening can be considered as the use of a test when the prior probability is very low, a situation that will naturally lead to potential problems with false-positive findings. The distinction between a screening test and a screening programme is also important: the screening programme includes a definition of who should be screened at what frequency, plus systems for bringing individuals to screening, making a definitive diagnosis in those that screen positive, and treating those with the target disorder. It is important to note that screening may lead to overdiagnosis, the identification of disease that will not affect an individual's health, either because the individual dies first, or because the disease process is so indolent as to be clinically innocent over meaningful time horizons.

A systematic consideration of the components of a screening programme and its evaluation will be important. Important points include the following:

- The target of the screening: illness in the preclinical detection phase. The availability of an effective treatment—one which leads to better outcomes if used in the preclinical phase than later—is a prerequisite for a sensible screening programme.
- The screening test itself. In the screening situation (typically low prevalence of the target lesion), the tension between sensitivity and specificity will be an important issue, often with no crisp resolution. To avoid false-positive results, a screening test should have a low false-positive rate (high specificity). In longer, or more advanced courses, the difficulties of defining the sensitivity and specificity of a screening test can be discussed (Morrison 1992).
- The expected pattern of events after screening. Early detection may simply advance the time of diagnosis without altering the final outcome (lead time bias). This almost inevitable improvement in observed case survival will force consideration of selection into screening, lead time biases, and length-biased sampling.
- The effectiveness of screening. Some students may assume early detection will necessarily be beneficial and even necessarily save money. The potentially negative impact of screening on those who falsely screen negative or those who falsely screen positive will need to be illustrated, as will the possible harms from ineffective treatment. This discussion can lead naturally to cost-effectiveness analysis and expected value decision techniques.

In advanced courses, the observational assessment of screening interventions can be discussed, as well as the computation of the operating characteristics of screening tests. Case-control investigation of screening (Morrison 1992; Cronin et al. 1998) is a particularly challenging issue that involves many quantitative topics.

Discussion of prevention requires only a few additional concepts, including the distinction between primary, secondary, and tertiary prevention. Otherwise, much of the basics of prevention can recapitulate analogous discussions of screening regarding target population and effectiveness. The students will probably enter with biases regarding the inevitable benefits of prevention, and arguments similar to those used for screening can be used to discuss these. On a population level, the fact is that most cases of many diseases develop in individuals at low or moderate risk. The high-risk individuals with the most to gain from prevention contribute little to the population burden of the target disease. This phenomenon points to the 'prevention paradox', that is, that interventions beneficial on a population level may not benefit most individuals (i.e. those at low or moderate risk; Rose 1981).

Expected value decision analysis and economic analyses

Expected value decision analysis (Sox et al. 2013) is often used to aid expert panels in recommending clinical policies and can in principle be the basis for decision aids for clinical decision-making. It has value in teaching a systematic approach to decisions when the outcome is uncertain, and so may foster clear thinking even if the technique is not explicitly used clinically. Expected value decision analysis is prescriptive (showing how decisions could rationally be made), rather than descriptive (showing how they are actually made). The approach is highly modelled and begins with explicit enumeration of the possible actions the clinician can take and the possible health state outcomes of those actions (e.g. class IV congestive heart failure). The process rests on several assertions:

- The possible outcomes experienced after a decision depend on future unpredictable events. The probability of each potential outcome can be estimated as the product of the probabilities of the events leading to the outcome.
- One can represent the value of a health state by a number (e.g. length of life, utility); the expected value of an outcome is its probability multiplied by its value.
- One may compare decision options according to their expected values (i.e. the sum of the expected values of the outcomes of each decision option).

Figure 24.1 shows how the terms described in these assertions fit together in a decision tree.

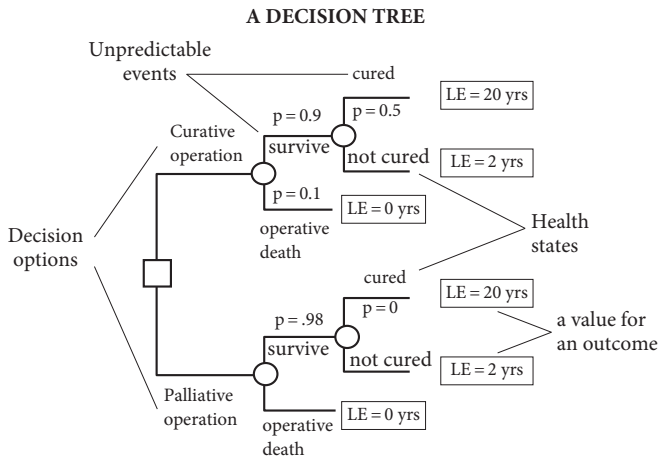


Fig. 24.1 A hypothetical decision tree.

p = probability, LE = life expectancy.

Economic analyses incorporate cost issues into expected value decision analyses. The basic idea is to calculate expected outcomes for two measures: cost and value. New concepts that are involved include discounting, direct versus indirect benefits, the perspective of the analysis, and the role of marginal (incremental) analyses. Cost-effectiveness analysis is more commonly encountered than cost-benefit analyses and merits correspondingly greater emphasis. Both are measures of the efficiency with which resources are converted into health outcomes. The basic principle is to compare health-care interventions on the basis of the cost per unit of outcome. The most commonly used clinical measure of outcome is quality-adjusted life years (QALYs), which is equivalent to years of life in good health. Cost-effectiveness analysis is always comparative. The basic relationship is:

$$\text{cost-effectiveness} = \left(\text{cost}_1 - \text{cost}_2\right)/\left(\text{QALY}_1 - \text{QALY}_2\right),$$

where the subscripts refer to the compared interventions. Thus, the cost-effectiveness of intervention 1 versus intervention 2 is the ratio of the added expected cost of the intervention 2 to the gains (or loss) in QALYs due to intervention 2 relative to intervention 1. An important point of emphasis is that cost-effectiveness analysis does not indicate whether a practice is worth doing; it simply assesses which intervention is the most efficient use of the money.

Decision analysis is a natural entry to sensitivity analysis, an approach that is useful in many research settings. Here, the expected value of each decision option is calculated after substituting different values for one of the model

parameters. If the same decision option has the highest expected value for all values of the parameter, the decision analysis is robust with respect to that parameter.

Molecular and genetic clinical epidemiology

Molecular and genetic methods are being increasingly used in clinical medicine for diagnosis, screening, risk stratification, and choice of targeted treatments. The application of these measurements is generally accommodated well in the framework of diagnosis, prognosis, etc., developed above. However, some of the details of this application can provide interesting issues for discussion of those basic modalities; for example,

- genetic variants of uncertain significance are an example an imperfect gold standard;
- gene penetrance is essentially the prognosis of the genetic variant; thus, studies conducted in genetic clinics bring issues of selection bias;
- very sensitive molecular tests for cancer are likely to introduce stage migration and identify pseudocancer; and
- proteomic and gene expression data have high dimensionality (a large number of predictor variables) and so are prone to overfitting (falsely optimistic assessment of their predictive value).

Teaching method and format: assessment of students

Clinicians who attend a formal classroom course may be borrowing clinical time to attend. Beeper interruptions tend to be common, and students may not be able to attend all sessions because of responsibilities for patient care. The optimum is to require that students turn off beepers and attend virtually all sessions. The reality is that this may be impossible. As a consequence, building some redundancy into formal courses for these students is useful: material presented in class can be obtained with readings, etc.

Once the readings have been assigned and the theory is presented, application of the concepts to clinical problems should be stressed. As a preliminary, however, simple exercises that review the basic concepts might be introduced, including simple true/false questions about the concepts, etc. The bulk of the exercises could then include clinical scenarios or critique of articles.

Discussion of the limitations of 'naive' (non-quantitative) approaches to clinical problems is important for three reasons: to reinforce the quantitative concepts themselves, to highlight the clinical relevance of the discipline, and to help

motivate students from purely clinical backgrounds. Examples of these discussion points include errors or confusion engendered by

- reliance on numerator data alone for conclusions that require risks or rates;
- confusion between relative and attributable risks;
- confusion between statistical and clinical significance;
- failure to anticipate a low positive predictive value in the setting of low disease prevalence;
- reliance on a poorly validated clinical prediction rule; and
- choice of a treatment option that has low expected utility.

In many ways, teaching clinical epidemiology solely in classroom courses removes the discipline from its natural context. An ideal is the incorporation of clinical epidemiology into routine clinical teaching. A complete basic methods course is probably not feasible in the context of the usual menu of medical conferences. Nonetheless, with an active and suitably trained faculty, clinical epidemiological concepts can be reinforced during the usual medical conferences and teaching rounds.

Evaluation of learning in a clinical epidemiology course can proceed through a variety of paths. Particularly useful are critiques of articles dealing with the concepts covered, or solutions to clinical scenarios. Advertisements in medical journals often provide interesting and timely examples of material for interpretation in these exercises.

Teaching resources for clinical epidemiology

McMaster University maintains an active programme regarding the teaching of evidence-based medicine and offers periodic courses for teachers (<http://ebm.mcmaster.ca/>). In addition, several media sources can provide relevant teaching materials. The American College of Physicians (ACP) Journal Club provides summaries and sophisticated interpretations of recent articles (<http://www.acpjc.org/> (fee required)). With suitable methodological critiques, these can be used as teaching examples, or (without the critiques) as exercises.

Several websites have been developed that incorporate potentially useful material. These include the following:

- the 'Supercourse' website (<http://www.pitt.edu/~super1/>), which contains complete lectures (with slides), some of which address topics in clinical epidemiology;
- the website of the Oxford Centre for Evidence-Based Medicine (<http://www.cebm.net/>), which contains online summaries of many clinical epidemiology topics, as well as some useful examples;

- the epidemiolog.net website (<http://epidemiolog.net/>), which contains links to online texts, courses, and other teaching resources; and
- the Cochran Collaboration website (<http://www.cochrane.org/>), which has modules for training, plus information regarding hundreds of formal meta-analyses.

In addition to the books quoted in this chapter, beginning students in clinical epidemiology might find the following books useful:

- Guyatt, G. H., Rennie, D., Meade, M. O., and Cook, D. J. (2008) *Users' Guides to the Medical Literature* (2nd edn). New York: McGraw Hill. This book contains reprints of a series of articles in the *Journal of the American Medical Association* regarding critical appraisal of the medical literature, and the practice of evidence-based medicine.
- Fletcher, R. H., Fletcher, S. W., and Fletcher, G. S. (2012) *Clinical Epidemiology: The Essentials* (5th edn). Philadelphia, PA: Lippincott Williams & Wilkins. A general text on clinical epidemiology.
- Haynes, R. B., Sackett, D. L., Guyatt, H. H., and Tugwell, P. (2006) *Clinical Epidemiology: How To Do Clinical Practice Research*. Philadelphia, PA: Lippincott Williams & Wilkins. A general text on clinical epidemiology, with an emphasis the conduct of studies.
- Straus, S. E., Richardson, W. S., Glasziou, P, and Haynes, R. P. (2005) *Evidence-Based Medicine: How To Practice and teach EBM*. Edinburgh: Elsevier/Churchill Livingstone.

Conclusion

Since clinical epidemiology involves research design and statistical analysis, it could incorporate much of the field of epidemiology in general. This breadth is one of the challenges that confront teachers in this field. Arguably, the best resolution is to recognize that a clinical epidemiology course is not primarily a course in epidemiological study design; focus on the important clinical topics (prognosis, diagnosis, treatment, and clinical trials) will also be important.

The backgrounds of the students may leave the instructor with additional challenges: some courses may involve both clinicians who have little quantitative background and epidemiologists or statisticians with almost no clinical experience. In this situation, it may be particularly difficult to keep the different types of students challenged and interested, and not overwhelmed or bored. A focus on the clinical part of clinical epidemiology will help motivate clinical students while the technical concepts appropriate for students with quantitative backgrounds are discussed.

References

Alderman, M. H. and Lamport, B. (1990) Labelling of hypertensives: a review of the data. *Journal of Clinical Epidemiology*, **43**: 195–200.

Barnett, A. G., van der Pols, J. C., and Dobson, A. J. (2005) Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, **34**: 215–20.

Baron, J. A. (1994) Uncertainty in Bayes. *Medical Decision Making*, **14**: 46–51.

Black, W. C. and Welch, H. G. (1993) Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *New England Journal of Medicine*, **328**: 1237–43.

Bland, J. M. and Altman, D. G. (1995) Comparing two methods of clinical measurement: a personal history. *International Journal of Epidemiology*, **24** Suppl. 1: S7–14.

Cronin, K. A., Weed, D. L., Connorm, R. J., and Prorok, P. C. (1998) Case-control studies of cancer screening: theory and practice. *Journal of the National Cancer Institute*, **90**: 498–504.

Elmore, J. G. and Feinstein, A. R. (1992) A bibliography of publications on observer variability (final installment). *Journal of Clinical Epidemiology*, **45**: 567–80.

Feinstein, A. R., Sosin, D. M., and Wells, C. K. (1985) The Will Rogers phenomenon: stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *New England Journal of Medicine*, **312**: 1604–8.

Hemingway, H. (2006) Prognosis research: why is Dr. Lydgate still waiting? *Journal of Clinical Epidemiology*, **59**: 1229–38.

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., and Smith, G. D. (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, **27**: 1133–63.

Machin, D. and Fayers, P. M. (2010) *Randomized Clinical Trials*. Chichester: Wiley-Blackwell.

Maclure, M. and Willett, W.C. (1987) Misinterpretation and misuse of the kappa statistic. *American Journal of Epidemiology*, 126: 161–9.

McDowell, I. (2006) *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press.

Morgenstern, H., Kleinbaum, D. G., and Kupper, L. L. (1980) Measures of disease incidence used in epidemiologic research. *International Journal of Epidemiology*, **9**: 97–104.

Morrison, A. S. (1992) *Screening in Chronic Disease*. New York: Oxford University Press.

Pepe, M. S., Janes, H., Longton, G., Leisenring, W., and Newcomb, P. (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology*, **159**: 882–90.

Rose, G. (1981) Strategy of prevention: lessons from cardiovascular disease. *British Medical Journal* (*Clinical Research Edition*), **282**: 1847–51.

Sorensen, H. T., Lash, T. L., and Rothman, K. J. (2006) Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology*, **44**: 1075–82.

Sox, H. C., Higgins, M. C., and Owens, D. K. (2013) *Medical Decision Making* (2nd edn). Oxford: Wiley-Blackwell.

Streiner, D. and Norman, G. (2008) *Health Measurement Scales: A Practical Guide to their Development and Use* (4th edn). New York: Oxford University Press.

Taylor, D. C. (1979) The components of sickness: diseases, illnesses, and predicaments. *Lancet*, **2**: 1008–10.

Walker, A. M. (1996) Confounding by indication. *Epidemiology*, **7**: 335–6.

Wasson, J. H., Sox, H. C., Neff, R. K., and Goldman, L. (1985) Clinical prediction rules: applications and methodological standards. *New England Journal of Medicine*. **313**: 793–9.

Wright, J. C. and Weinstein, M. C. (1998) Gains in life expectancy from medical interventions—standardizing data on outcomes. *New England Journal of Medicine*, **339**: 380–6.

Chapter 25

Study of clustering and outbreaks

Paul Elliott and Anna Hansell

Syllabus

1. Introduction to the study of disease clusters and clustering.
2. Cluster investigation and statistical appraisal.
3. Interpretation, potential confounding, and bias.
4. Introduction to the practicals and case studies.
5. Practical/case study 1: anophthalmia in England and Wales.
6. Practical/case study 2: estimation of the SMR.
7. Practical/case study 3: denominator issues.
8. Practical/case study 4: follow-up study using routine data.
9. Group presentations and discussion.
10. Assessment.
11. Summary and lessons learnt.

Introduction to studying clustering and outbreaks

This chapter is concerned with understanding the principles underlying the detection and investigation of possible clusters of disease, and the extent to which they might contribute to knowledge on potential environmental causes. The focus is on the investigation of outbreaks of *chronic* disease, or at least outbreaks potentially related to chronic exposures to environmental pollutants. The investigation of *acute* outbreaks, including of infectious diseases (such as food poisoning), in general requires a different approach.

The issues covered in this chapter would usually be considered part of postgraduate training in epidemiology or public health. However, the topic has general interest and would be suitable for teaching to an undergraduate audience with a quantitative background and an interest in health and the environment. Accounts of the area are given in Elliott and Wartenberg (2004) and Elliott and

Savitz (2008), and general considerations are covered in public health guidelines for cluster investigation, such as that from the Centers for Disease Control and Prevention (1990).

Teaching objectives

The teaching objectives are summarized as follows.

- Describe the context within which claims of disease clusters arise.
- Explain what is meant by a disease 'cluster' so that the students are able to define it in statistical or causal terms.
- Explain the distinction between a single disease cluster and a tendency for disease 'clustering'.
- Explain the importance of random variability in helping to determine spatial and temporal patterns of disease, especially involving small populations and small numbers of cases.
- Ensure that the students are aware of the data issues that affect interpretation of possible disease clusters, including those associated with the cases (numerator), the population at risk (denominator), and geographic linkages between the two.
- Explain the concept of the standardized mortality or morbidity ratio (SMR) and describe how this is often used in the initial assessment of a possible disease cluster.
- Stress the importance of potential confounding, especially that due to socioeconomic factors.
- Describe the nature of and potential for bias in disease cluster studies.
- Describe what steps are required in a cluster investigation and when further investigation may be appropriate.

Teaching content

In line with the above objectives, the course should cover the following topics:

1. context of the reporting of disease clusters;
2. definitions of a disease cluster;
3. random variability;
4. initial investigation and statistical appraisal; and
5. interpretation, potential confounding, and bias.

Each of these is now discussed in turn.

Context of the reporting of disease clusters

It is important that students are aware of the key trends in disease occurrence over the last century so that they can put the reporting of disease clusters in proper context. The declines in the major infectious diseases, and increases (and later declines) in coronary heart disease and lung cancer in the UK, US, and other developed countries are largely related to improvements in hygiene and standards of living, and more recently to lifestyle, diet, and smoking patterns. These might be characterized as the *internal* environment, while public attention has focused increasingly on the possible health impact of the *external* or *physical* environment. This includes ionizing or non-ionizing radiation, emissions from local industry, or perceived threats from contaminated land, the water supply, or waste disposal.

These issues can usefully be discussed with reference to the WHO's Global Burden of Disease Project, which looked at the attributable mortality and disability adjusted life years (DALYs) in 1990 and 2010 for sixty-seven risk factors. The project participants estimated that the five leading risk factors are high blood pressure, tobacco smoking, household air pollution, a diet low in fruits, and alcohol use (see Fig. 3 in Lim et al. 2012). Quantified attributable risk fractions differ depending on the disease being considered. The students should be invited to make their own ranking before comparison is made with those from the Global Burden of Disease Project.

Public concern is often accompanied by suspicions of disease excess in the local community—so-called cluster reports. These are often picked up by the media; examples can readily be found to illustrate the point.

Definitions of a disease cluster

Clearly, it is important to discuss with the students what is meant by the term 'disease cluster'. Knox (1989) suggested three alternative definitions (Box 25.1).

With the first of these, little weight can be placed on any particular local excess, as there is a general tendency for the disease itself to clump over space. An example is Hodgkin's disease (Alexander et al. 1989), implicating a possible infectious etiology for that condition.

The second of Knox's definitions is purely statistical and does not deal with the question of possible etiology but is the approach taken in surveillance. The problem of how to set the boundaries that determine 'sufficient size and concentration' needs to be discussed with the students: if the threshold is set too low (i.e. high sensitivity but low specificity), large numbers of 'false-positive' clusters are bound to occur. Neutra (1990) addresses these issues (see 'Group discussion and debate: Rothman versus Neutra'). Richardson et al. (2004) give

Box 25.1 Definitions of a disease cluster according to Knox

A cluster is/is not:

a geographically and/or temporally bounded group of occurrences—

of a disease already known to occur characteristically in clusters;

of sufficient size and concentration to be unlikely to have occurred by chance;

related to each other through some social or biological mechanism, or having a common relationship with some other event or circumstance.

both a useful summary of the statistical issues and worked examples with regard to sensitivity and specificity of disease mapping studies, which are closely associated with detection of disease clusters.

Knox himself preferred the third definition, as it relies on some notion of causation. He also posed the fundamental question underlying any cluster investigation: 'Is it real?'

Students should be made aware of a potential, well-recognized pitfall in cluster investigation: the 'Texas sharpshooter effect', which relates to choice of the geographical or temporal boundary of a cluster post hoc (i.e. after seeing the spatial and/or temporal distribution) so as to maximize the density of cases. This relates to a joke that the Texan sharpshooter first fires his bullet and then draws the target around the bullet hole (Olsen et al. 1996).

Random variability

A key issue for the students to understand is the nature of random variability. First, there is a vast (often unacknowledged) problem of multiple testing in cluster studies, since to arrive at a single apparent 'cluster' in a given area and time period, a large number of observations ('tests') are potentially made—over thousands of small areas, different time periods, disease categories, age and gender groups, and so on. Second, for rare diseases in small geographic areas, or over short time periods, the expected number of cases is low, leading to large random variability.

Clearly, the combination of multiple testing in one direction only (high values of the relative risk) and small expected numbers will lead to the reporting of

many 'false-positive' clusters. Students need to appreciate that distinguishing between false-positives and any true increase in risk ('Is it real?' in Knox's terminology) is at the heart of cluster detection and investigation.

Initial investigation and statistical appraisal

Protocols from the US Centers for Disease Control (1990) and Drijver and Melse (1992), as well as other protocols available on the European Surveillance of Congenital Anomalies website (<http://www.eurocat-network.eu/clusterinvestigationprotocols>) give good accounts of the approach to a cluster investigation and are useful as a basis for group discussion. It is important that the students are aware of the steps in a cluster investigation and have an understanding of the key issues. The steps (see Elliott and Wakefield 2001) include the following:

1. form a hypothesis;
2. select the geographic region(s) and time period(s) for study;
3. assemble and check the data;
4. obtain an estimate of risk;
5. write a report and feed back the results; and then
6. end the study or pursue further investigation, as indicated.

Most guidelines also cover the development of a communication strategy with the communities and people affected at an early stage of the investigation.
Key points for discussion include

- differentiating between a priori and post hoc observations (the latter are notoriously difficult to evaluate as statistical testing is formally invalidated);
- the importance of data checking, as many apparent clusters may simply reflect anomalies in the data (e.g. see 'Denominator issues in the study of clustering and outbreaks: Dalgety Bay in Scotland' for a case study where census estimates in an area of rapid growth gave biased risk estimates);
- the case definition used, which may need review as the investigation proceeds (a specific rather than a sensitive definition is likely to be the most useful);
- the problem of 'boundary shrinkage' which may accentuate estimates of disease risk (according to Olsen et al. (1996), the more narrowly the underlying population is defined, the fewer will be the number of expected cases, the greater will be the estimate of the excess rate, and often the more pronounced will be the statistical significance);
- calculation of a risk estimate (for example, the SMR), which is often obtained by comparing the observed numbers of cases in areas near and more distant

from a point source (or in high and low pollution areas), with the numbers expected based on the age and gender distribution of the population, using indirect standardization (Aylin et al. (1999), in an appendix, give details of the calculation of the SMR in the context of disease clusters and small area disease mapping, whereas Breslow and Day (1987) discuss the properties of the SMR more generally; see 'Estimation of the SMR in the study of clustering and outbreaks' for a worked example);

- specialized statistical tests for clustering are available (e.g. Moran's I, the Cuzick–Edwards test, Stone's test, and spatial scan statistics (for which SaTScan freeware is available; <http://www.satscan.org>), although these are generally not indicated in initial investigations of cluster reports (in addition, a significant statistical test alone cannot 'prove' the presence or absence of a cluster and different tests may give different answers);
- the importance of potential socio-economic confounding, which can distort estimates of disease risk (Jolley et al. 1992);
- the need to manage hopes and expectations among the local community (which may be unrealistically high); and
- when to proceed with or terminate an investigation (Centers for Disease Control and Prevention 1990).

Interpretation, potential confounding, and bias

Students need to be aware of the difficult issues of interpretation if an apparent disease excess in a particular area is found. In particular, there are many sources of potential bias: Elliott and Wakefield (2000) provide a discussion on this. Issues to be considered include

- confounding, particularly associated with socio-economic factors affecting disease rates for small areas;
- bias in the identification of areas for study, since an informal and poorly defined selection process is undertaken (areas at apparently low risk are never selected!);
- the 'boundary shrinkage' problem, leading to accentuated estimates of risk;
- low population numbers and small numbers of cases, giving imprecise estimates;
- ill-defined hypotheses or, in some cases, lack of hypothesis;
- poorly defined or absent exposure data;
- incomplete or inaccurate health and population data, and migration between areas, especially where latency periods are prolonged (most cancers); and

- reporting bias, for example, if medical professionals are more likely to diagnose cases given concerns about a particular industrial installation or other putative exposure source.

Teaching method and format

This should be a combination of lecture/seminar format, small-group teaching, case studies, worked examples, and private study. A suggested breakdown of time would be lecture/seminar, 30 per cent; case studies and worked examples (small-group teaching), 30 per cent; and reading/preparation time and private study, 40 per cent. Seminars would typically include twenty to thirty students; small groups (for the case studies) might include three or four students working together and reporting back to the main group.

Typically, the course will be held over two or three days (four to six sessions, depending on the number of case studies), although it may be longer if computing sessions are included.

Some worked examples and case studies are given below. These could be set as class exercises, used as the basis for discussion, or modified for use in assessment. They do not require computing facilities, although the calculation of the SMR example requires calculators. Similar exercises using computer programs could be added, particularly for more advanced courses. An example is the GeoBUGS software for disease mapping, which is public domain software (<http://www.mrc-bsu.cam.ac.uk/software/bugs/thebugs-project-geobugs/>). For mapping and calculation of the SMR around a point source, the UK Small Area Health Statistics Unit has developed the Rapid Inquiry Facility software in partnership with the US Centers for Disease Control and Prevention (Beale et al. 2010); this software is also in the public domain (<http://www.sahsu.org/content/rapid-inquiry-facility>), although it is currently (2013–14) in redevelopment.

Worked examples and case studies

Cluster report: anophthalmia in England and Wales

This case study illustrates the media reporting of alleged clusters of a rare disease and the steps that were required to investigate the clusters. In *The Observer* newspaper of 17 January 1993, it was claimed under the headline 'Mystery of babies with no eyes' that there were localized clusters of anophthalmia across England and Wales. This is a serious birth defect where the infant is born without eye tissue (either unilateral or bilateral); it is closely related to microphthalmia, where some eye tissue is present. The newspaper linked the 'clusters' to the use of the pesticide Benomyl.

The article was accompanied by a map which illustrated well many of the problems and pitfalls in the reporting of alleged clusters. First, it was unclear from the map and accompanying article (and subsequent enquiry) what was the source of the cases, whether or not they were truly anophthalmia, whether or not there was duplicate reporting, and so forth. Second, no timescale was given for the occurrence of the cases, and only in Louth in the north-east of England was any clear geographic boundary given (a forty-mile radius), so that it would have been impossible to determine a denominator for the alleged clusters and hence obtain an estimate of the risk. Without such basic information, it was not possible to make any judgement as to whether or not there was any true excess of this condition, nor of course whether there was any geographic association with pesticide use (Dolk and Elliott 1993).

As a result of the concerns raised by this and other newspaper reports, a birth prevalence study of anophthalmia and microphthalmia was set up. No evidence of specific clusters or generalized clustering was found, although there did appear to be an unexplained excess in rural compared with urban areas (Dolk et al. 1998).

Estimation of the SMR

This is a hypothetical worked example based on the anophthalmia case study. The scenario is that a search of the national congenital anomaly register indicates six cases of anophthalmia in the study area in the last ten years and eighty-nine cases in the region of England which includes the study area. Table 25.1 presents numbers of cases and population at risk (number of live births) stratified by maternal age—as incidence increases with maternal age. Column (d)

Table 25.1 Study and reference population cases and births by maternal age.

Study area				**Region**	
Maternal age in years (a)	**Number of births (b)**	**Number of cases (c)**	**Expected number of cases (d)**	**Number of births (e)**	**Number of cases (f)**
<20	1,465	0	0.10	42,920	3
20–24	5,012	0	0.30	134,068	8
25–29	5,988	1	0.50	189,832	16
30–34	6,336	2	0.96	185,074	28
35–39	3,588	2	0.85	109,382	26
>40	895	1	0.28	25,594	8
Total	23,285	6	3.02	686,870	89

should be left blank for the students, so that they can calculate expected numbers.

Part 1. Ask students to calculate the SMR. Students should first calculate the expected number of cases in each maternal age group (column (d)) and then sum them to obtain total expected number of cases in the study area. In order to calculate expected cases in each maternal age group, the rate from the reference population (the region) should be applied to the study population (i.e. column (f)) divided by column (e) and multiplied by column (b). The SMR can be readily calculated as the ratio of observed to expected cases multiplied by 100; confidence intervals can be obtained using an online calculator such as that suggested in Part 2.

Part 2. Ask the students to calculate the probability that the number of observed cases in an area is larger than the expected number by chance with an online tool such as the EUROCAT online calculator (the tool requires the numbers of observed cases and expected cases; <http://www.eurocat-network.eu/pagecontent.aspx?tree=Asimplestatisticaltest>. This gives the *P*-value, in this case, $P = 0.084$ (i.e. is not statistically significant). The example should lead to a discussion about how a doubling of a small number of cases can still be a chance finding. Students should also appreciate that a *P*-value alone cannot 'prove' the presence or absence of a cluster.

Denominator issues: Dalgety Bay in Scotland

This is an example of a small area investigation of cancer incidence in a small town in east-central Scotland following the discovery of small amounts of radium-226 on the foreshore (Black et al. 1994). This was thought to emanate from radium-based luminous paint that was used in military aircraft that were disposed of by burning during the 1940s. Because of recent in-migration, the demography of Dalgety Bay had changed during the study period, so that careful attention to the denominator was needed in the calculation of SMRs. The paper also demonstrates, in its Table 2, the effect of allowing for socio-economic deprivation in the calculation of SMRs (observed:expected ratios). Since Dalgety Bay is an affluent area, some risk estimates (e.g. those for lung cancer and stomach cancer) were higher following adjustment for deprivation, while others were lower (especially skin cancer, where the observed:expected ratio was 1.50 with a 95% confidence interval (CI) ranging from 1.05 to 2.08 with age and sex adjustment, and 1.38 with a 95% CI ranging from 0.97 to 1.91 with additional adjustment for deprivation). The students should be invited to discuss these issues, as well as their views on the conduct and interpretation of the study.

Follow-up study using routine data: cancer risk in childhood and mobile phone base stations in Great Britain

This study illustrates how initial cluster reports, public concern, and scientific uncertainties led to a national investigation using routine cancer data. Reports of apparent cancer clusters near mobile phone base stations were difficult to interpret because of small numbers and possible selection and reporting bias. A case-control study was initiated to examine associations between early childhood cancers and distance/estimated radiofrequency exposures during pregnancy from macrocell base stations in Great Britain, using national cancer registry data to avoid selection bias, and post hoc analyses (Elliott et al. 2010). The analysis examined distance/modelled exposures, 1996–2001, for 1,397 children aged 0–4 years who developed cancer in 1999–2001, and 5,588 controls from the birth register. No association was seen between cancer risk in young children and living near or estimated radio frequency exposures during pregnancy from a mobile phone base station.

Group discussion and debate: Rothman versus Neutra

In a landmark paper, Rothman (1990) argued that cluster investigation was hardly ever indicated; instead, he advocated a traditional hypothesis-led investigation, and (as required) the remediation of environmental concerns per se, without reference to the health statistics. Neutra (1990), in his reply, largely agreed with Rothman's analysis. Although he found some justification for cluster investigation, this was only for rare diseases, with large relative risks (>20) associated with high specificity exposures and where case aggregation met the criterion for minimum size (>5). These two papers together give an excellent background for the interested student.

Assessing students' achievements

Assessment will need to be based on the requirements of the course. Possibilities include short-answer questions to interpret findings of apparent disease 'excess' in a particular locality/time period, and/or to delineate steps in the investigation. Students might be asked to carry out a paper critique of a cluster investigation or to present in small groups to the rest of the class. Another possibility is to carry out a computational exercise (e.g. calculation of the SMR and its assessment), using either a computer or calculator-based practical. In all cases it will be essential to determine that the students have gained the critical appraisal skills and understanding needed to allow them to properly evaluate the cluster literature.

References

Alexander, F. E., Williams, J., McKinney, P. A., Ricketts, T. J., and Cartwright, R. A. (1989) A specialist leukaemia/lymphoma registry in the UK. Part 2: Clustering of Hodgkin's disease. *British Journal of Cancer*, **60**: 948–52.

Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., Jarup, L., Arnold, R., Wheeler, G., and Elliott, P. (1999) A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit. *Journal of Public Health Medicine*, **21**: 289–98.

Beale, L., Hodgson, S., Abellan, J. J., Lefevre, S., Jarup, L. (2010) Evaluation of spatial relationships between health and the environment: the Rapid Inquiry Facility. *Environmental Health Perspectives*, **118**: 1306–12.

Black, R. J., Sharp, L., Finlayson, A. R., Harkness, E. F. (1994) Cancer incidence in a population potentially exposed to radium-226 at Dalgety Bay, Scotland. *British Journal of Cancer*, **69**: 140–3.

Breslow, N. E. and Day, N. E. (1987) Statistical Methods in Cancer Research. Volume II: *The Design and Analysis of Cohort Studies*. Lyon: International Agency for Research on Cancer.

Centers for Disease Control and Prevention. (1990) *Guidelines For Investigating Clusters of Health Events*. <http://www.cdc.gov/mmwr/preview/mmwrhtml/00001797.htm>, accessed 29 June 2013.

Dolk, H., Busby, A., Armstrong, B. G., and Walls, P. H. (1998) Geographical variation in anophthalmia and microphthalmia in England, 1988–94. *British Medical Journal*, **317**: 905–10.

Dolk, H. and Elliott, P. (1993) Evidence for 'clusters of anophthalmia' is thin. *British Medical Journal*, **307**: 203.

Drijver, M. and Melse, J. M. (1992) *Protocol 1*. http://www.eurocat-network.eu/content/Clusters-Protocol-1.pdf>, accessed 30 October 2014.

Elliott, P. and Savitz, D. A. (2008) Design issues in small-area studies of environment and health. *Environmental Health Perspectives*, **116**: 1098–104.

Elliott, P., Toledano, M. B., Bennett, J., Beale, L., de Hoogh, K., Best, N., and Briggs, D. J. (2010) Mobile phone base stations and early childhood cancers: case-control study. *British Medical Journal*, **340**: c3077.

Elliott, P. and Wakefield, J. C. (2000) 'Bias and confounding in spatial epidemiology', in P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, eds, *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press, pp. 68–84.

Elliott, P. and Wakefield, J. C. (2001) Disease clusters: should they be investigated, and if so, when and how? *Journal of the Royal Statistical Society: Series A* (*Statistics in Society*), **164**: 3–12.

Elliott, P. and Wartenberg, D. (2004) Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives*, **112**: 998–1006.

Jolley, D., Jarman, B., and Elliott, P. (1992) 'Socio-economic confounding', in P. Elliott, J. Cuzick, D. English, and R. Stern, eds, *Geographical and Environmental Epidemiology: Methods for Small-area Studies*. Oxford: Oxford University Press, pp. 115–24.

Knox, E. G. (1989) 'Detection of disease clusters', in P. Elliott, ed., *Methodology of Enquiries into Disease Clustering*. London: Small Area Health Statistics Unit, pp. 17–20.

Lim, S. S. et al. (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**: 2224–60.

Neutra, R. (1990) Counterpoint from a cluster buster. *American Journal of Epidemiology*, **132**: 1–8.

Olsen, S. F., Martuzzi, M., and Elliott, P. (1996) Cluster analysis and disease mapping—why, when and how? *British Medical Journal*, **313**: 863–6.

Richardson, S., Thomson, A., Best, N., and Elliott, P. (2004) Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, **112**: 1016–25.

Rothman, K. J. (1990) A sobering start for the cluster busters' conference. *American Journal of Epidemiology*, **132**: S6–13.

Chapter 26

Medical databases

Henrik Toft Sørensen and John A. Baron

Introduction to medical databases

Medical database is a term that refers to all types of registries and databases that contain health-related data. The cost, complexity, and risk of selection bias often associated with primary data collection such as randomized clinical trials has led to an increasing use of disease registries and databases as an alternative or supplementary data source for studies of many epidemiological and clinical questions (Schneeweiss 2006).

Existing databases are increasingly used and are often the only feasible source with which to examine delayed health effects. They contain information on large study populations and can provide information on exposures, disease outcomes, and some potential confounding factors for subjects in those populations (Sørensen et al. 1996; Schneeweiss and Avorn 2005). Thus, medical databases can be very useful for the efficient study of etiologic associations in large populations and for the evaluation of the utilization, effectiveness, and safety of medical interventions in clinical settings.

Health registries are databases that include patients who are diagnosed with a specific disease, or who undergo a particular procedure. In contrast, *administrative databases* select individuals according to their residency in a specified geographic area, membership in a health insurance programme, or attendance at a particular hospital or clinic—irrespective of the diseases the patients have or the procedures used to treat them. Many administrative databases have more comprehensive data than classic health registries.

The classic *health* or disease *registry* will thus typically include all cases of a particular health condition or health-related exposure in some defined population. The *register* is the actual listing of the cases; in formal terms, the registry is the system of ongoing registration (and possibly follow-up), together with the register (Porta et al. 2008). Conceptually, a register can just be a collection of records but in practice it involves a formal database. Of note, however, the distinction between registries and registers is not universally accepted and the terms are frequently interchangeable in the international literature. Data regarding

post-diagnosis or post-procedure events are a component of many classic registries and add considerably to the value of the data in them. In some countries, notably those in Scandinavia, disease/procedure registries are increasingly integrated with administrative databases, further encouraging database research.

A major aim of many disease registries is the monitoring of disease occurrence over time (as in cancer and birth registries). Trends in disease occurrence and comparisons between different populations have generated important hypotheses on the causes of these diseases and evaluation of the effect of prevention programmes. Consequently, the reporting of the relevant data to the registry is typically mandated by the government sponsoring the registry. Although the term *registry* is commonly used to refer to the associated register, the system of registration is an important determinant of the quality of the data and consequently also of the research conducted with them. This will need to be stressed in teaching about research using such data. Similar considerations apply to administrative databases. In North America there has been a substantial growth in the use of databases from insurance plans for medical research.

Classic health registries have operated in the Nordic countries for many years. The first known population-based disease registry was established as early as 1856: the Leprosy Registry in Norway. At the beginning of the twentieth century, registries for causes of death, cancer incidence, and births sprouted in Scandinavia and the UK. Examples from North American are the National Registry of Myocardial Infarction (Gurwitz et al. 1998) and cancer registries such as those sponsored by many US states and Canadian provinces and those in the US Surveillance Epidemiology and End Results programme.

Technological developments and increased demand for documentation of activities related to health care have led to a considerable increase in the number of medical databases during recent decades. The term *clinical database* refers to a subgroup of administrative databases that contain clinical information regarding patients' medical histories, laboratory and radiological findings, and treatments. The data are collected during the course of routine clinical care but are also often used to assess the effectiveness of treatments and adverse effects and to address other research questions. A special group of clinical databases are the *quality assessment databases*, for example, those in Scandinavia, as they have very detailed clinical data.

For several reasons, the number of database studies is likely to increase in coming years. On the one hand, many research questions can be addressed quickly and efficiently with these available data sources. On the other hand, other avenues of research are becoming more difficult: many epidemiological studies are reporting declining participation rates, privacy restrictions on data use are becoming more burdensome, and clinical trials are becoming more expensive.

Teaching objectives

Teaching about the use of medical databases should give students insight into

1. examples of important medical databases available for research;
2. the three key elements that characterize medical databases;
3. the strengths and limitations of databases as data sources;
4. the assessment of data quality;
5. other aspects of database research; and
6. teaching method and format.

Existing medical databases available for research

Published studies using medical databases are an excellent introduction for students, since they immediately illustrate the potential of database research. If possible, the use of local databases should be featured; Box 26.1 contains some examples of database studies from North America and Europe.

Three key elements that characterize a medical database

To be able to conduct valid database-based research, it is important for students to understand the likely advantages, disadvantages, and problems associated with that work. Regardless of the purpose of the existing database, the methodological, problems facing the researcher who uses the data are similar (Sørensen et al. 1996). A first step is understanding the components of the database (Baron and Weiderpass 2000).

1. The population. Each database has a referent population—the individuals whose medical experience is described by its data. This population may consist of individuals residing in certain geographic areas, or members of an insurance plan, or patients seen at certain medical facilities. In population-based databases, enumeration of the underlying population can provide denominators for rate computations and the basis for an unbiased selection of controls in a population-based case-control study. In Denmark, for example, the entire population is recorded in the Danish Civil Registration System (Frank 2000), allowing population-based research regarding the whole country.
   a. Studies based on geographic areas are often relatively straightforward. For example, a study of the association between use of venous thromboembolism and the risk of an arterial cardiovascular event (Sørensen et al. 2007) illustrates the population-based design and an obvious corresponding population-based comparison cohort.

Box 26.1 Examples of various types of medical databases used in epidemiology

1. Classic medical health registries
   - Cancer incidence and mortality data (Luo et al. 2008)
   - Birth registry data (Mahon et al. 2007)
   - Multiple sclerosis data (Koch-Henriksen et al. 2011)
2. Administrative databases
   - Hospital databases (Ludvigsson et al. 2011)
   - Medicare and Medicaid databases from the US (Cooper et al. 2006; Setoguchi et al. 2007)
   - Kaiser Permanente Medical Care databases from the US (Mahadevan et al. 2007)
   - Health care databases from Canada (Curkendall et al. 2006; Xu et al. 2006)
   - General practitioners' medical record databases from the UK and Denmark (Huerta et al. 2007; Paulsen et al. 2012)
3. Clinical databases
   - The Danish Breast Cancer Collaborative Group Database (Blichert-Toft et al. 2008)
   - The Duke Cardiovascular Disease Databank (Whellan et al. 2006)
   - The Danish National Stroke Database (Palnum et al. 2008)
   - Northern New England Cardiovascular Disease Registry (Dacey et al. 2007)
   - Swedish Coronary Angiography and Angioplasty Registry (SCAAR) (James et al. 2009)

References

Blichert-Toft, M., Christiansen, P., and Mouridsen, H. T. (2008) Danish Breast Cancer Cooperative Group–DBCG: History, organization, and status of scientific achievements at 30-year anniversary. *Acta Oncologica*, **47**: 497–505.

Cooper, W. O., Hernandez-Diaz, S., Arbogast, P. G., Dudley, J. A., Dyer, S., Gideon, P. S., Hall, K., and Ray, W. A. (2006) Major congenital malformations after first-trimester exposure to ACE inhibitors. *New England Journal of Medicine*, **354**: 2443–51.

Curkendall, S. M., Lanes, S., de Luise, C., Stang, M. R., Jones, J. K., She, D., and Goehring, E. Jr. (2006) Chronic obstructive pulmonary disease severity and cardiovascular outcomes. *European Journal of Epidemiology*, **21**: 803–13.

Box 26.1 Examples of various types of medical databases used in epidemiology *(continued)*

Dacey, L. J. et al. (2007) Long-term survival after surgery versus percutaneous intervention in octogenarians with multivessel coronary disease. *Annals of Thoracic Surgery*, **84**: 1904–11.

Huerta, C., Johansson, S., Wallander, M. A., and Rodríguez, L. A. G. (2007) Risk factors and short-term mortality of venous thromboembolism diagnosed in the primary care setting in the United Kingdom. *Archives of Internal Medicine*, **167**: 935–43.

James, S. K., Stenestrand, U., Lindbäck. J., Carlsson, J., Scherstén, F., Nilsson, T., Wallentin, L., and Lagerqvist, B. (2009) Long-term safety and efficacy of drug-eluting versus bare-metal stents in Sweden. *New England Journal of Medicine*, **360**: 1933–45.

Koch-Henriksen, N., Stenager, E., and Brønnum-Hansen, H. (2011) Studies based on the Danish multiple sclerosis registry. *Scandinavian Journal of Public Health*, **39**: 180–4.

Ludvigsson, J. F., Andersson, E., Ekbom, A., Feychting, M., Kim, J. L., Reuterwall, C., Heurgren, M., and Olausson, P. O. (2011) External review and validation of the Swedish national inpatient register. *BMC Public Health*, **11**: 450.

Luo, J., Adami, H. O., Reilly, M., Ekbom, A., Nordenvall, C., and Ye, W. (2008) Interpreting trends of pancreatic cancer incidence and mortality: a nation-wide study in Sweden (1960–2003). *Cancer Causes and Control*, **1**: 89–96.

Mahadevan, U., Sandborn, W. J., Li, D. K., Hakimian, S., Kane, S., and Corley, D. A. (2007) Pregnancy outcomes in women with inflammatory bowel disease: a large community-based study from Northern California. *Gastroenterology*, **133**: 1106–12.

Mahon, B. E., Ehrenstein, V., Nørgaard, M., Pedersen, L., Rothman, K. J., and Sørensen, H. T. (2007) Perinatal risk factors for hospitalization for pneumococcal disease in childhood: a population-based cohort study. *Pediatrics*, **119**: e804–12.

Palnum, K. D., Petersen, P., Sørensen, H. T., Ingeman, A., Mainz, J., Bartels, P., and Johnsen, S. P. (2008) Older patients with acute stroke in Denmark: quality of care and short-term mortality. A nationwide follow-up study. *Age and Ageing*, **37**: 90–5.

Paulsen, M. S., Andersen, M., Thomsen, J. L., Schroll, H., Larsen, P. V., Lykkegaard, J., Jacobsen, I. A., Larsen, M. L., Christensen, B., and Sondergaard, J. (2012) Multimorbidity and blood pressure control in 37 651 hypertensive patients from Danish general practice. *Journal of the American Heart Association*, **31**: e004531.

Setoguchi, S., Glynn, R. J., Avorn, J., Levin, R., and Winkelmayer, W. C. (2007) Ten-year trends of cardiovascular drug use after myocardial infarction among community-dwelling persons ≥65 years of age. *American Journal of Cardiology*, **100**: 1061–7.

Whellan, D. J., Tuttle, R. H., Velazquez, E. J., Shaw, L. K., Jollis, J. G., Ellis, W., O'Connor, C. M., Califf, R. M., and Borges-Neto, S. (2006) Predicting significant coronary artery disease in patients with left ventricular dysfunction. *American Heart Journal*, **152**: 340–7.

Xu, W., Tamim, H., Shapiro, S., Stang, M. R., and Collet, J. P. (2006) Use of antidepressants and risk of colorectal cancer: a nested case-control study. *Lancet Oncology*, **7**: 301–8.

2. The medical events, diseases, and other data. All databases are 'abstracts' in the sense that they only include selected data on conditions, events, or other outcomes and contain only basic demographic information. Few databases will contain the detailed clinical data that are typically seen in medical records or questionnaire-based studies. Most administrative databases do not contain data regarding lifestyle habits and personal characteristics such as cigarette smoking, alcohol use, BMI, etc. In most databases, numeric coding systems are used to characterize the medical events or diagnoses covered, so the descriptions may be limited by details of the coding system. For example, in the ninth version of the WHO's *International Classification of Diseases* (ICD-9), laterality in paired organs like the legs is not recorded. Changes in coding systems may introduce lack of comparability and detail. An example is that venous thrombosis in the upper extremities had a separate code in the eighth version of the ICD (ICD-8) but not in the tenth version (ICD-10). The codes for the different versions of the international disease classifications can be found on the following web pages: <http://www.wolfbane.com/icd/icd8.htm>, <http://www.cdc.gov/nchs/icd/icd9.htm>, and <http://www.who.int/classifications/icd/en/>.
3. The organization of data collection and management. In databases many individuals enter data; consequently, there is the risk of introducing substantial variation and error, especially when the data elements are not well defined. This is a particular problem in databases in which 'free text' is often used, as, for example, in the description of imaging and of pathological specimens. The completeness and validity of the reporting of many diagnoses and procedures will depend on the administrative mandates involved and the structure of the database with compulsory fields. In some databases such as billing-derived databases, lack of reporting involves a financial penalty. In many clinical or administrative databases there may be little motivation to record covariates in hospital or clinical visits, and the accuracy of database information regarding even main diagnoses and procedures may depend on how the clinical record is translated into data points in the database.

The strengths and limitations of databases as data sources

The widespread use of medical databases in research is a reflection of their many advantages (Sørensen et al. 1996; Baron and Weiderpass 2000; Schneeweiss 2006; Sørensen et al. 2006). Primarily because the data already exist in the database, the time spent on the study is likely to be considerably less than that spent on studies using primary data collection, greatly increasing efficiency and decreasing costs. The large size of many databases offers the potential for precise estimates of effect

and the possibility of studying rare exposures or outcomes. Large database studies are often the only possibilities to do meaningful subgroup analyses.

There may be several other advantages of database research:

1. population-based databases, that are essentially complete, greatly reduce the risk of selection bias;
2. the collection of registry data independently of the research project minimizes the risk of certain types of bias (e.g. those associated with recall, nonresponse, and/or impact on the diagnostic process of the attention caused by the research itself); and
3. databases with long follow-up have the ability to provide information on effects that have long induction period, or on long-term prognosis.

Nonetheless, there are many limitations to database research. An important basic point to convey to students is that even though database research need not be 'quick and dirty', it is still amenable to error (Schneeweiss and Avorn 2005; Sørensen et al. 2006). Database research is no different from other types of research in the need for careful attention to methodology and sample size. In addition, it is important for students to understand that the research topic needs to suit the database and that many questions cannot be addressed through this sort of study. Research questions that involve lifestyle factors, for example, are usually impossible to address in databases because the relevant data are simply not available, except, for example, in the General Practice Research Database in Britain. Even in this case, it is not entirely clear when or how well the lifestyle factors were recorded over a long lifespan. Issues that involve very detailed clinical information may also not be adequately addressed with database research, except for special clinical databases.

Other limitations may not be recognized by students. A common problem with studies that use databases is the variable degree of completeness and validity of the information and the possible lack of information on potential confounders and clinical detail (e.g. disease severity, location, treatment, etc.; Sørensen et al. 1996; Baron and Weiderpass 2000). The limitations are largely a result of data availability, since the methods of data collection are predetermined, not controlled by the researcher, and sometimes very difficult or impossible to validate. As noted above, data required for the research question may not have been collected at all (Sørensen 1997). Important examples are detailed clinical data such as disease severity or results of laboratory and radiographic tests. An important problem is that, very often, the indication for a specific treatment is not clearly recorded, and the student should be familiar with the concept of 'confounding by indication'. In addition, difficulty with coding is an important potential limitation of databases.

Poor data quality can constitute an insurmountable obstacle to a database study. Missing data for data fields nominally collected are another potential limitation to database research. Obviously, this may occur when compulsory fields are not completed but it is also an issue when the reporting requirements for a database are not clear. For example, the need to record diagnoses in hospital discharge list may be variously interpreted, leaving some diagnoses unreported. An absent required data point will be immediately evident: by definition, the data field will contain a missing value code. In contrast, underreporting of diagnoses in discharge records is difficult to quantify. Inaccurate or missing data generally tend to bias associations toward the null hypothesis rather than to cause spurious associations, as long as they occur in equal proportion in the groups to be compared (Sørensen et al. 1996). Nonetheless, missing or incomplete data in medical databases may be an important source of bias in models based on such data.

Assessment of data quality

The student must be familiar with the basic concepts of bias, sensitivity, specificity, predictive value, and precision. Students will need to have had at least an introductory course in epidemiology to be familiar with basic epidemiological terminology and study designs before taking a specific course in the use of databases.

Completeness of population coverage

Some cases will be omitted from even the best database system through data errors, failure to record the relevant diagnosis or procedure, or other administrative errors such as lost records, etc. The completeness of a database can be estimated by the proportion of true cases in the target population for the study question which is correctly classified in the database. Evaluation of completeness can be accomplished in two general ways: through *individual* or *aggregated* assessments. In the former, cases in the database are individually identified in one or more independent complete reference sources and the sensitivity estimated (Goldberg et al. 1980). Some researchers have compared different independent data sources, and the numbers of missing cases have been estimated in a capture–recapture model (Sørensen et al. 1996). If the underlying population is well defined, an estimate of the specificity can also be obtained (Sørensen et al. 1996). The completeness of database coverage may also be assessed by aggregation methods in which the total number of cases in the data source is compared with the total number in other sources, or the expected number of cases is calculated by applying rates from demographically similar populations or by simulation (which uses the information system to simulate patterns of incomplete

reporting to examine the possible effect on a specific dependent variable; Goldberg et al. 1980).

Data quality

Data quality problems can be derived from errors in the dataset from incorrect data entry or lack of available information, or from inaccuracies in the original source of information. Unfortunately, it is usually impossible to identify the latter sort of problem, despite medical record review.

Students are often confused by the many different terms that are used to describe validation. *Validity* is the extent to which a variable measures what it is intended to measure. Lack of validity is referred to as *bias* (Rothman et al. 2008). If this terminology is used, the two most often used validity measures are sensitivity and a predictive value of inclusion in the database. The completeness of a database is largely determined by the sensitivity of the registration process that created it. The completeness of the persons in the cohort may not be as important as the completeness of the registered data, including the degree of missing data. The demands for completeness and representativeness depend on the research question. For several analytical studies the degree of completeness may be less important than whether the misclassification is random or differential (Sørensen et al. 1996). A comparison between two data sources alone does not provide the opportunity to estimate the specificity, unless one of the sources is known to be complete.

Record review is often used to validate diagnoses in hospital discharge, birth, cancer, or demographic data in exiting databases. Most often, the ratio between the number of correctly registered persons (e.g. registered cases who actually had meningococcal disease) and all registered persons (e.g. all registered as having meningococcal disease) is measured. This is essentially the *positive predictive value of registration*. It is important to stress that these sorts of statistics reflect only one dimension of validity, namely, the predictive value of the registrations, and not the completeness of data. In particular, sensitivity and specificity are not directly assessed.

Furthermore, the data quality evaluation should also provide detailed insight into the extent of missing data of other variables. For each binary single variable, it should be considered whether a negative value means that exposure or outcome has taken place, or whether the data are simply missing. For example, in the Danish Birth Registry, smoking status can be coded 'yes', 'no', or 'missing'. We have found that an absent data point most likely represents a 'no'. For ordinary and continuous variables such as weight, blood pressure, and staging, it is only possible to check the validity through review of the medical records.

In a more advanced course, statistical methods on how to deal with *missing data* can be covered. When few data are missing, little harm will be done but, if

many data are missing, this may create bias, unless the missing data are completely at random. The concept of *imputation* can be reviewed, a technique in which missing data are replaced by plausible values predicted from that individual's available data (Schneeweiss 2006). In more advanced courses, the concepts of sensitivity analysis and external adjustment for unmeasured confounders and the basic principles of bias analysis would be covered (Lash et al. 2009).

Other aspects of database research

It is essential for the student to understand the importance of the number of persons and outcomes included in the database used in a particular study. It may also be relevant to know the distribution of the various variables and especially the number of outcome events, since this information is likely to be important in designing studies. For rare diseases such as for instance idiopathic thrombocytopenic purpura (ITP), one Nordic country is not large enough to do a safety study on the treatments of ITP. Meaningful subgroup analyses may also require very large datasets.

In longitudinal studies, information concerning the registration period(s) is essential for a database study to take into account *induction periods* and *latent periods* for etiologic research and the follow-up period for prognostic studies (Sørensen et al. 1996). The induction period is the time required for a specific cause to lead to the incidence of a disease process (whether symptomatic or not); the latent period is the time between disease initiation and its clinical manifestation. Data sources with registration periods of a few years will seldom be suitable for etiologic research regarding chronic diseases that have prolonged induction periods. Since most databases cover a limited time period, the relevant exposures or outcomes may be uncertain, unless the database collected information covering time periods considerably longer than the induction and latent periods of the disease process under study.

Some cases, for example, may not become manifest until several years after the disease is incident: congenital heart diseases are classic examples. As a result, research that examines the association of pregnancy characteristics with congenital heart disease will overlook less severe cases, unless a search is made for those that are diagnosed years after birth. This is an example of *right censoring*.

Databases, by definition, cover events only during a defined time interval. Events occurring before registration began cannot be included, nor can events that occur afterward. If events are used as exposures in an observational analysis, then follow-up commencing soon after the beginning of registration will be

open to incomplete exposure information. This is an example of *left truncation of the data*. In more advanced courses it will often be relevant to introduce the concept of new-user design in studies of treatments (Ray 2003). The design tries to handle this problem.

The problem of events that might occur later than the data covered by the registry (right censoring) can be handled by standard techniques of survival analysis. Beginning students may not be familiar with right censoring but, if time permits, an introduction of the relevant concepts of survival analysis can be included in the course.

Another type of data absence involved observation 'holes' within the time period of the registry. For example, prescription databases do not cover drug exposure during hospitalizations and over-the-counter medication.

A problem related to the large volume of available data in databases is the risk of significant associations that emerge by chance. Because of the ease of conducting some database analyses, a large number of exposures and/or outcomes can be assessed, leaving the analysis open to chance findings. Another important problem related to the large size of databases is that relatively weak associations may, nonetheless, still be statistically significant because of the large number of end points included in the analysis. Of course, very small relative risks—even if statistically significant—may be easily generated by relatively subtle biases, so these findings will require especial scrutiny with regard to confounding, selection and ascertainment biases, etc. In addition, when even small relative risks are statistically significant, there is an acute need to differentiate between statistically significant associations and those that are clinically relevant. Application of absolute risk and risk difference (used to compute the number needed to prevent/treat and harm) may be more useful for determining clinical significance than use of relative risk estimates (Fletcher et al. 2014). A number of medical registries are easily accessed, and different research groups may use the same data for analysing identical or similar issues.

Older computerized records may be formatted or structured in such a way that their use is difficult for research. For example, the diagnostic categories may be outdated or age categorized only into very wide bands. The investigator must critically review existing documentation to assess the appropriateness of the data for their intended use. If such documentation does not exist, protocols, record layout and codes, data entry instructions, published material, analyses, and technical reports will need to be surveyed and appropriate studies of completeness and validity carried out, all with respect to the specific context of the study. Lack of experience with programming or lack of collaboration with a programmer may cause potential errors in preparing data for analysis or in the analysis.

There are several types of problems that may limit the usefulness of coded diagnoses in databases (Steinberg et al. 1990; Sørensen 1997):

- variation in coding between persons doing the coding;
- errors in coding of the recorded data;
- lack of coding of certain data points, especially of co-morbidities and life-style factors;
- inherent limitations in the specificity of available codes; and
- errors and variation in the clinical diagnoses on which the coding is based.

Furthermore, codes, and even the layout of records, may change periodically. Changes in diagnostic criteria, classifications, and methods (e.g. the recent change from ICD-9-CM to the ICD-10 disease classification system) frequently cause problems in comparisons of data over longer periods (Sørensen et al. 1996; Baron and Weiderpass 2000). Furthermore, the diagnostic criteria for a number of diseases such as acute myocardial infarction (AMI) are changing; when teaching, we often use an example from a *BMJ* article that shows that the incidence of AMI started increasing in 2002 because of changes in diagnostic criteria, even though the incidence of AMI has been steadily declining for the past twenty-five years (Schmidt et al. 2012).

Since database research may be considerably strengthened through the use of data from multiple sources, the linkage of subjects across the different databases may be crucial (Sørensen 1997). Record linkage techniques can help to identify the same person in different files, so that a more complete medical 'picture' may be obtained, perhaps including inpatient visits, cancer registry data, and vital status, for example. Linkage is accomplished via a personal identification number. However, a complete high-quality record linkage may not always be possible. The best population-based data sources are probably the extensive data-linkage networks in Scandinavia, where each person is assigned a unique personal registration number at birth, allowing record linkage between several independent data systems and vital statistical registers (also called external record linkage; Sørensen et al. 1996; Baron and Weiderpass 2000; Frank 2000). We often use an example on statins and fractures, for which two studies were carried out based on the same data from the GP database in the UK (Meier et al. 2000; Van Staa et al. 2001). This possibility should be kept in mind when carrying out meta-analyses in which the same patients can be included several times.

Teaching method and format

The optimum introductory course consists of lectures, simple exercises, and reading of papers. Basic epidemiological concepts should be kept in mind at all

times as issues more specific to databases are considered. Group exercises may include examining articles dealing with the same research question in which primary data collections or registers have been used. Examples of articles that illustrate problems with database research may be particularly valuable since they serve to caution students about the need for careful work, while they illustrate particular technical points. An example is a study of breast implants and breast cancer risk. Berkel and colleagues (1992) reported a reduced risk of breast cancer subsequent to breast augmentation. A number of potential errors in the study were later noted, and a reanalysis could not confirm the findings. Exercises such as these give the students a sense of the strengths and weaknesses of the various research approaches. Other examples should include problems with inconsistent use of epidemiologic concepts.

We avoid the use of computers when teaching introductory courses, except basic calculations on a pocket calculator. This is partly because these types of courses deal with the understanding of basic principles, and partly because the students often do not have the necessary background for detailed computations. However, even simple examples of estimation on a pocket calculator can often be of value. On more advanced courses it is relevant to use computer exercises but analyses of dynamic cohorts, for example, demand an extensive and sophisticated programming. There is hardly any room for this during class, except when one is teaching experienced epidemiologists, as the programming problems will overshadow the aim of the teaching. However, exercises of analyses of simple case-control studies are often very valuable. In a more advanced course, analyses involving actual medical databases would be appropriate.

Assessing students' achievements

A useful evaluative exercise is asking students to evaluate papers: their critiques of database research will provide valuable insight into the students' mastery of the topic. We often give them papers on the same topic based on primary and secondary data collection and ask them to identify strengths, limitations, and bias problems. We have sometimes used papers regarding venous thromboembolism and subsequent cancer risk, a topic that is covered by both primary and registry data (Prandoni et al. 1992; Baron et al. 1998; Sørensen et al. 1998). When teaching we often also use examples of randomized studies on the same topic (i.e. the association between post-menopausal hormone-replacement therapy and the risk of AMI). It can be an advantage for this evaluation exercise to delete the discussion sections of the papers and ask the students to write the discussion.

References

Baron, J. A., Gridley, G., Weiderpass, E., Nyren, O., and Linet. M. (1998) Venous thromboembolism and cancer. *Lancet*, **351**: 1077–80.

Baron, J. A. and Weiderpass, E. (2000) An introduction to epidemiological research with medical databases. *Annals of Epidemiology*, **10**: 200–4.

Berkel, H., Birdsell, D. C., and Jenkins, H. (1992) Breast augmentation: a risk factor for breast cancer*? New England Journal of Medicine*, **326**: 1649–53.

Fletcher, R. H., Fletcher, S. W., and Fletcher, G. S. (2014) *Clinical Epidemiology: The Essentials* (5th edn). Philadelphia, PA: Lippincott Williams & Wilkins.

Frank, L. (2000) When an entire country is a cohort. *Science*, **287**: 2398–9.

Goldberg, J., Gelfand, H. M., and Levy, P. S. (1980) Registry evaluation methods: a review and case study. *Epidemiological Review*, **2**: 210–20.

Gurwitz, J. H., Gore, J. M., Goldberg, R. J., Barron, H. V., Breen, T., Rundle, A. C., Sloan, M. A., French, W., and Rogers, W. J. (1998) Risk for intracranial hemorrhage after tissue plasminogen activator treatment for acute myocardial infarction. *Annals of Internal Medicine*, **129**: 597–604.

Lash, T. L., Fox, M. P., and Fink, A. K. (2009) *Applying Quantitative Bias Analysis to Epidemiologic Data (Statistics for Biology and Health)*. Dordrecht: Springer.

Meier, C. R., Schlienger, R. G., Kraenzlin, M. E., Schlegel, B., and Jick, H. (2000) HMG-CoA reductase inhibitors and the risk of fractures. *Journal of the American Medical Association*, **283**: 3205–10.

Porta, M., ed., Greenland, S. and Last, J. M., associate eds. (2008) *A Dictionary of Epidemiology. A Handbook Sponsored by the IEA* (5th ed.). Oxford: Oxford University Press.

Prandoni, P., Lensing, A. W., Büller, H. R., Cogo, A., Prins, M. H., Cattelan, A. M., Cuppini, S., Noventa, F., and ten Cate, J. W. (1992) Deep-vein thrombosis and the incidence of subsequent symptomatic cancer. *New England Journal of Medicine*, **327**: 1128–33.

Ray, W. A. (2003) Evaluating medication effects outside of clinical trials: new-user designs. *American Journal of Epidemiology*, **158**: 915–20.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008) *Modern Epidemiology* (3rd edn). Philadelphia: Lippincott Williams & Wilkins.

Schmidt, M., Jacobsen, J. B., Lash, T. L., Bøtker, H. E., and Sørensen, H. T. (2012) 25 year trends in first time hospitalisation for acute myocardial infarction, subsequent short and long term mortality, and the prognostic impact of sex and comorbidity: a Danish nationwide cohort study. *British Medical Journal*, **344**: e356.

Schneeweiss, S. (2006) Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and Drug Safety*, **15**: 291–303.

Schneeweiss, S. and Avorn, J. (2005) A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology*, **58**: 323–37.

Sørensen, H. T. (1997) Regional administrative health registers as a resource in clinical epidemiology: a study of options, strengths, limitations and data quality provided with examples of use. *International Journal of Risk and Safety in Medicine*, **10**: 1–22.

Sørensen, H. T., Horvath-Puho, E., Pedersen, L., Baron, J. A., and Prandoni, P. (2007) Venous thromboembolism and subsequent hospitalisation due to acute arterial cardiovascular events: a 20-year cohort study. *Lancet*, **370**: 1773–9.

Sørensen, H. T., Lash, T. L., and Rothman, K. J. (2006) Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology*, **44**: 1075–82.

Sørensen, H. T., Mellemkjær, L., Steffensen, F. H., Olsen, J. H., and Nielsen, G. L. (1998) The risk of a diagnosis of cancer after primary deep venous thrombosis and pulmonary embolism. *New England Journal of Medicine*, **338**: 1169–73.

Sørensen, H. T., Sabroe, S., and Olsen, J. (1996) A framework for evaluation of secondary data sources for epidemiological research. *International Journal of Epidemiology*, **25**: 435–42.

Steinberg, E. P., Whittle, J., and Anderson, G. F. (1990) Impact of claims data research on clinical practice. *International Journal of Technological Assessment of Health Care*, **6**: 282–7.

van Staa, T. P., Wegman, S., de Vries, F., Leufkens, B., and Cooper, C. (2001) Use of statins and risk of fractures. *Journal of the American Medical Association*, **285**: 1850–5.

Chapter 27

Teaching epidemiology inside and outside the classroom

J. H. Abramson

Introduction to teaching epidemiology inside and outside the classroom

There is no single ideal way to teach epidemiology. Teaching takes place in different situations, and its techniques and content differ. A good teaching programme is one that is geared to its students' needs, capacity, interests, and preferences and that utilizes the available situations and techniques to provide learning opportunities that will achieve its objectives.

This chapter reviews some features of the teaching of epidemiology inside and outside the classroom. It starts with discussions of teaching objectives and other factors that affect the choice of teaching methods, and then deals in turn with conventional classroom methods, laboratory teaching (problem-solving and other exercises), self-instruction, problem-oriented projects, distance learning, and combined methods of teaching. Separate consideration is then given to teaching in the hospital and in the field (with special attention to teaching in a community health centre).

Teaching objectives

The educational objectives, or learning outcomes—that is, what it is hoped the students will learn to do—require careful specification. They may be categorized broadly as follows:

1. *Knowing* epidemiology: the ability to state specified facts about epidemiological terms, principles, and methods, and specified epidemiological study findings, for example, facts about the epidemiology of selected diseases.
2. *Using* epidemiology: the capacity to interpret and appraise epidemiological findings and to apply epidemiological knowledge when performing activities that are not primarily epidemiological. Medical students may be expected to acquire skill in making clinical decisions founded on evidence

about the prevalence and incidence of diseases, the variability of biological findings, the prognostic significance of risk factors and risk markers, the reliability and validity of test results, and the results of clinical trials and analytic observational studies. Students who are or will be involved in community health care, public health practice, or policy formulation may similarly be expected to acquire skill in the use of epidemiological evidence when making managerial or policy decisions. For many teachers, the main goal is that students should be able to 'think logically' when appraising causal relationships. The capacity to read articles critically may be specified as an objective.

3. Practising (*doing*) epidemiology: the capacity to perform specified epidemiological activities. These may range from simple tasks such as calculating rates, through more complex activities—such as questionnaire design, the maintenance of disease registers, and data processing and analysis—to the comprehensive planning and performance of studies. Emphasis may be put on the ability to perform research designed to yield generalizable conclusions or on more pragmatic studies aimed at the improvement of health or health care in a specific community or population.
4. An ability to *teach* epidemiology, especially to other members of the health team, is often specified as an additional objective. A related but usually neglected objective is the ability to *communicate* epidemiological findings—to the public and the media, and especially to decision-makers—in such a way as to maximize their use and minimize their misuse.

The choice of situations and methods for teaching depends on the educational objectives, and vice versa. Students will learn to 'use' and 'practise' epidemiology best if they are taught in a context in which epidemiology is 'used' and 'practised', and if they are given practical experience (with suitable supervision) in 'using' and 'practising' it.

In general, the educational objectives should be adapted to the needs of the students and the requirements of the community in which the students will work. Objectives may vary for medical students, physicians in clinical practice, clinicians who wish to undertake research, nursing students, nurses and other professionals engaged in patient care, medical assistants and other auxiliaries, aspirant field epidemiologists and other professional epidemiologists, public health students and practitioners, community physicians, health administrators, health service researchers, statisticians and auxiliary personnel engaged in epidemiological and health service research, and so on. A school of public health may offer basic courses, additional courses for students pursuing a master's degree in public health and other students who need additional training in

the use of epidemiology and interpreting and communicating epidemiologic findings, and advanced courses for students who intend to perform epidemiologic research, preferably coordinated with classes in biostatistics (Gange 2008). For students who are not committed to a career in a health profession but are taught epidemiology as a part of general or liberal undergraduate education, it has been suggested that it should be taught broadly as a way of thinking, with applications not restricted to the health field; detailed educational objectives have been proposed for such courses (Riegelman et al. 2010). Teaching of epidemiology to schoolchildren has been advocated to improve their reasoning and research skills and enhance their ability to analyse and solve complex problems; materials and suggestions for the teaching of epidemiology to schoolchildren are available at <http://www.cdc.gov/excite/classroom.html>.

There is no consensus about educational objectives for specific categories of students, even in a single country, and no single 'best' teaching programme. Some medical schools are satisfied if students learn the epidemiology of common diseases, or the epidemiological foundations for disease prevention or health promotion. Some focus on applications in clinical decision-making. Other schools emphasize critical reading; and some expect students to learn how to make a community diagnosis and evaluate a community health programme. Although most teachers of epidemiology would probably regard all these aims as important, educational objectives are often limited by the restrictive views of curriculum committees, or other constraints. Moreover, specific objectives will differ in communities with different health problems.

The variation in objectives means that no teaching programme can be regarded as the single best one for universal application to a given category of student.

There are at least three obvious themes that should pervade all epidemiology teaching. First, things may not be what they seem—findings can be deceptive and can be influenced by random variation, selection bias, misclassification, and other forms of information bias, by confounders, and by incorrect reporting. Second, association does not necessarily mean causation. And third, epidemiological findings, whether descriptive or analytic, should be applied to the control of real-world health problems. In the spirit of 'translational' epidemiology (Khouri et al. 2010) ('the role of epidemiology in translating scientific discoveries to population health impact'), students should be brought in contact not only with studies of the distribution and determinants of health-related states and events but also with studies of the efficacy of diagnostic methods, with clinical and programme trials, and with studies of real-world impacts on population health.

Emphasis should be given to the value of epidemiological findings (descriptive or analytical) in community diagnosis and in the planning, monitoring, and evaluation of health care. But the students should also learn why it is inadvisable to rely on unreplicated epidemiological findings. Epidemiological findings should always be 'taken with care' (Abramson 1997). It is arguable that all students should be exposed to some of the issues raised by Ioannidis (2005) in his paper 'Why most published research findings are false' and acquire at least a nodding acquaintance with the Bayesian principles that may affect interpretation.

Training in sophisticated epidemiological fields such as molecular pathological epidemiology (Ogino et al. 2012) and genomic epidemiology (Khouri et al. 2004) and in new analytical techniques such as propensity score modelling (Williamson et al. 2011) should probably be confined to selected future researchers.

Decisions about teaching methods

The choice of methods will be influenced not only by the educational objectives of the teaching programme, but also by the students' interests and preferences, and by practical considerations.

Students' interests and preferences

Some students have little interest in epidemiology—notably many medical students who see little relevance to their central concern, the care of individual patients. Apparently, it is only in the later years of the course of study, or after graduation, that there is a tendency towards a greater appreciation of the value of epidemiology (Moffat et al. 2004).

For such students, particularly in medical schools where teachers of other disciplines disparage epidemiology, it is important to concentrate on epidemiological topics of obvious clinical relevance, especially for the treatment of hospital patients.

The current interest in evidence-based medicine and evidence-based public health can motivate students to learn basic epidemiological principles and methods, and increasing use is being made of evidence-based medicine as a vehicle for the teaching of epidemiology. In Boston a controlled study indicated that exposure to four sessions on evidence-based medicine was enough to have a significant positive impact on critical appraisal skills (Ghali et al. 2000). A handbook on the practice and teaching of evidence-based medicine (Sackett et al. 2000) provides 'teaching tips' for motivating learning. These centre on the presentation of a clinical scenario (real or simulated), the selection of good journal articles that are relevant to this scenario, and 'closure' by reaching conclusions.

Even in schools where the climate for the teaching of epidemiology is unfavourable, efforts should be made to broaden the interests of students beyond the care of single patients. It needs only a few students to become interested in community health care or epidemiological research to make these efforts worthwhile, or at least to seem so, to a teacher who may otherwise find his/her work frustrating.

Students whose central interest is in health care at a group or community level, such as graduate students of public health or health administration, generally recognize the importance of epidemiology. But here, too, the teaching should be brought as close as possible to the students' interests in terms of specific situations (e.g. hospital or community health services), applications (research vs pragmatic epidemiology), and subject matter (heart disease, child health, etc.). A flexible programme that permits choices by students is of especial value. This may be achieved by giving students a choice of assignments or projects, and by providing a range of elective courses.

Students' preferences concerning methods of teaching should also be considered. Those who have had little experience, during or since their schooldays, of heuristic teaching aimed at encouraging them to discover for themselves rather than learning what they are told, may not take kindly to teaching based on a problem-solving approach. The capacity to solve problems is of central importance in 'practising' and 'using' epidemiology, and problem-solving exercises cannot easily be dispensed with; but they may be more acceptable and effective in some student groups than in others (in every group there will be some students who require more 'hand-holding' than others).

The effectiveness of all teaching methods varies for different students. Self-instruction may work well with some students and not with others. Some students enjoy working with computers more than others. A flexible programme that caters for different needs has obvious advantages. Students may be given a choice between attendance at lectures and reading, or between computerized and printed problem-solving exercises, and may be permitted to do the exercises in groups or alone, and in class or at home. Periodic surveillance of students' progress is especially important when such options are offered, as students may not always make the best choices. Not only the kind of teaching, but also the amount of teaching may vary for different students; special tutorials, summer projects, and elective courses have been advocated for those medical students who are interested in epidemiology, as well as very short courses for professionals outside the context of degree programmes (Bayona et al. 1994).

A number of evaluative studies—including randomized controlled trials and before–after studies, using both quantitative and qualitative methods—have demonstrated the value of specific teaching methods, such as a problem-based

approach (Dyke et al. 2001), discussion seminars (Marantz et al. 2003), an emphasis on the contribution of epidemiology to community-oriented health care (Öcek et al. 2008), and participation in epidemiological investigations (Soudarssaname et al. 1994). Useful though such studies certainly are, consideration should always be given to their generalizability—does the demonstrated success of a teaching programme for a specific type of student with specific teachers in a particular country necessarily mean that it will be equally successful elsewhere?

Practical considerations

It is hardly necessary to stress the importance of practical constraints that may affect teaching methods, such as the availability of time, teachers, and other resources, and the capacity and interests of teachers. If medical students are to be given 'hands-on' experience of the use of epidemiology in clinical work, for example, an obvious basic requirement is allotment of time for this during the clinical years.

Classroom teaching

Lectures

An exceptional lecture or series of lectures by an exceptional lecturer can be an inspiring experience. But by and large, lecturing is not the most effective way of teaching, although there may be little choice if teachers are scarce. Lectures are said to be as efficient as other methods with respect to the transfer of knowledge but less effective than discussions with respect to retention, transfer to new situations, problem-solving, thinking, and motivation for further learning (Krueger et al. 2004). In the UK, the proportion of public health teaching time allocated to lectures has decreased (Gillam and Bagade 2006). In a controlled trial of methods of teaching clinical epidemiology to medical students in Wales, lectures received the lowest evaluations (Gehlbach et al. 1985).

Good lectures can, of course, serve useful purposes. Apart from imparting facts, they can provide students with frameworks and models for use in their own explorations of the topic and they can sometimes communicate enthusiasm and a motivation to learn. The value of a lecture is enhanced by careful preparation, preparatory reading assignments for students, an interesting, audible, systematic, and not overlong presentation, appropriate use of the whiteboard or blackboard, computer presentations, slides, and handout material, sensitivity to the students' reactions, provision of opportunities for questions or discussions, and assignments based on the lecture. In a large class, student interest in the lecture can be stimulated by asking questions and requesting each student to discuss the answer with his or her neighbour (James et al. 2006). Repetition of

the lecture's main messages—but not too often, and preferably in different words—increases the chances that they will be remembered. 'Tell them what you're going to tell them, then tell them, and then tell them what you've told them' is a well-tried and effective tactic.

Lectures alone cannot easily be used to teach skills but, in conjunction with practical exercises, they can serve a valuable function by presenting systematic frameworks for the performance of activities (how to appraise a screening test, how to plan a survey, how to investigate an outbreak, how to control for confounding, how to do a meta-analysis, etc.).

Group discussions

Group discussions with active student participation (i.e. with interaction among students as well as with teachers) are more effective than lectures, particularly if the group is small.

Cognitive psychology, we are told

> shows that if we elaborate our learning by thinking or talking (explaining, summarizing, questioning) . . . we are more likely to remember it when needed later. Therefore, discussion is most appropriate: to help students learn to think; to assist students in the evaluation of the logic of evidence . . . ; to supply opportunities to apply principles; to increase awareness of problems and formulation of problems . . . ; to use the resources of the group for motivation (Krueger et al. 2004).

Discussions may centre on topics presented by teachers or students; some or all of the students may have been given reading assignments in advance. The teacher's roles include preventing the session from degenerating into a bad lecture by an unskilled student lecturer; drawing students into the discussion; guiding the discussion to ensure that the teaching potential of the topic is adequately exploited; and ensuring that students come away feeling they have learnt something, and knowing what it is that they have learnt. Billson (1986) offers practical advice on simple ways of nurturing constructive interaction in small groups.

'Case-discussion' sessions in which groups of twenty to twenty-four medical students discussed 'cases' after previously receiving and reading relevant materials were given strikingly positive evaluations (Marantz et al. 2003). These 'cases' were not patients but were doctors or medical students who had to make decisions or interpret information about, for example, the value of beta-carotene supplementation or the possible harmful effects of dietary sodium.

Discussion of students' presentations of articles—preferably, of articles that all students have been asked to read—provide an opportunity not only to learn about the topic but also to learn how to evaluate study methods and appraise the validity of findings and conclusions. They constitute a useful teaching method

in programmes where the ability to read critically is a defined objective, or where the purpose is to enable students to deal with evidence concerning specific clinical or public health issues. Some teachers like to use 'classical' papers (Buck et al. 1988) that illustrate important epidemiological concepts or methods; others prefer recent papers of current interest.

'Teach-back opportunities' may be extremely helpful (James et al. 2006)––for example, by asking small groups to learn about specific epidemiological study designs and then explain them to the whole class.

Group discussions play a central role in most problem-based learning programmes, in which students are required to gather, discuss, and use information relevant to a specific problem.

Discussions of recent articles may be organized in a 'journal club' framework, generally in the context of continuing education for physicians or other professionals. But a systematic review of controlled trials and before–after studies suggests that, for postgraduates, this kind of experience has little effect on skills or behaviour (Coomarasamy et al. 2003). This parallels the results of reviews of studies of instruction in critical appraisal (evidence-based medicine) skills, which showed improvements for undergraduates but gave limited evidence of effects for postgraduates (Norman and Shannon 1998; Hyde et al. 2000; Taylor et al. 2000).

Other methods

Role-playing may be a useful technique, for example, in teaching the use of interviews in surveys. Students interview one another in front of the class, using schedules they have prepared themselves, and this provides a basis for discussions of the effects of the wording or sequencing of questions, interview technique, and so on.

Other classroom methods include panel discussions, in which experts discuss a topic among themselves and answer students' questions, and symposia, in which presentations are made by a number of experts.

Laboratory teaching

'Laboratory' teaching is a convenient term for a variety of methods of involving students in practical activities that aim to teach epidemiological skills in a systematic way.

Problem-solving exercises

Problem-solving exercises generally present data derived from actual or imaginary studies and require students to analyse and appraise the facts, draw inferences, and reach decisions concerning the need for further investigations and/

or intervention. These activities are analogous to the 'hands-on' activities that have been successfully used in teaching clinical skills to medical students (Krueger et al. 2004). If the exercises relate to actual situations, their role in the teaching of epidemiology is similar to the role of the examination of patients in the teaching of clinical medicine (Lowe and Kostrzewski 1973). The exercises provide practical experience, under 'laboratory' conditions, in the use of epidemiological tools—methods of investigation, analysis, and data interpretation. Students may be required to gather data (generally from documentary sources), prepare tabulations and charts, perform statistical analyses, and formulate and test hypotheses. Library research may be needed. The exercises are often constructed serially (i.e. at each step students are asked what additional information they need and then given new facts and new tasks). The exercises may centre on epidemiological methods, the epidemiology of selected conditions, or specified epidemiological concepts. They are probably most effective if they deal with problems and techniques of obvious relevance to the students' interests.

The exercises may be carried out by students working alone or in small groups (of, say, four or five), inside or outside the classroom. Students are usually advised to summarize their answers in writing. When such exercises are done in a classroom, instructors may circulate among the groups to guide the students and review their answers, and the class may be brought together from time to time for a more general discussion.

Printed exercises designed to inculcate basic epidemiological concepts and skills, provided with solutions and comments, can be used either for self-instruction or in group situations and permit students or groups to work at their own pace. Exercises of this sort are provided by (inter alia) Abramson and Abramson (2001) and Hebel and McCarter (2012). Online exercises can be found at <http://www.epidemiolog.net>.

Advice on the planning and use of problem-solving exercises is provided by Noah (2001). The exercises can range from short ones, several of which can be covered in a single session, to an elaborate multistage investigation of a public health problem, requiring as much as one-twelfth of the total time of a master's programme in public health (Kreis et al. 1998). They may or may not require the use of a computer and of library resources.

Exercises in specific techniques

Exercises in specific techniques—performed in the classroom or as 'homework'—are commonly used to provide experience in performing epidemiological procedures and to reinforce or extend the student's knowledge of epidemiological principles or facts. Such exercises, which do not necessarily

present the intellectual challenge of problem-solving, can be used as adjuncts to lecture or reading courses. A course on research methods, for example, may include exercises on the choice of a study design, the formulation of study objectives, the choice of controls, the selection of a sampling method, the use of random numbers, the choice and definition of variables, the measurement of validity and reliability, the design of questions and questionnaires, data processing, and so on. These may be more than pencil-and-paper exercises: an unforgettable demonstration of observer variation, for example, can be staged by asking all the students in a group to measure a volunteer's height and weight, or to determine blood pressures based on a film of sphygmomanometer readings.

Some topics, for example, the use of statistical methods in epidemiology, can hardly be taught without the practical drill that systematic exercises provide.

Computer exercises

Students should be given experience in the use of computer software for epidemiological purposes. This may be done in the framework of problem-solving exercises, or whenever statistical issues come up in the course of group discussions, or via exercises specifically dedicated to computer procedures. Graduate students, in particular, should have experience in the use of computer databases and in computer analysis.

Various commercial statistical packages can be used for this purpose. Gange (2008), for example, recommends the use of Stata on the grounds that it is flexible, relatively inexpensive, and easy to use. Others prefer Minitab, on the grounds that novitiates 'mostly have little difficulty after only 3 hours of instruction' (Simpson 1995). But it is advisable to let students use readily available free programs, particularly in less affluent countries. An annotated list of statistical programs that can be downloaded from the Internet free or can be used online is provided by Abramson and Abramson (2008). Several lists of free or online statistical software are available on the Internet, notably, at <http://www.statpages.org> and <http://www.l-lists.com/en/lists/dz3a5t.html>.

Free statistical packages with a specifically epidemiological slant include EpiData (available at <http://www.epidata.dk>), Epi Info (<http://wwwn.cdc.gov/epiinfo>), OpenEpi (<http://www.openepi.com>) and WINPEPI (http://www.brixtonhealth.com/pepi4windows.html>). All four perform the basic statistical analyses required in most epidemiological studies, and the first two also help in the design of questionnaires or data entry forms and simplify the production of data sets. WINPEPI is a user-friendly 'Swiss army knife' package of programs for epidemiologists (Abramson 2004), and students usually find it easy to use, once they have learnt to focus on the specific modules and results that interest

them and to disregard the many others. There is no steep learning curve (or hardly any learning curve at all). WINPEPI can provide hands-on experience in all the procedures listed in a guide to the teaching of statistics in epidemiology (see chapter 5) and in all the statistical tools commonly used by clinical epidemiologists in patient care, as well as in many others that are less commonly used or not very easily found, such as the capture–recapture method. The package has considerable potential as a learning and teaching aid (Abramson 2011), both with respect to practical procedures in the planning and analysis of studies and with respect to important epidemiological concepts. It can, for example, demonstrate the effects of misclassification on prevalence rates and odds and risk ratios, assess inter-rater and intra-rater reliability, allow for the effects of multiple significance tests, appraise the influence of unmeasured confounders, illustrate the use of Bayes factors to appraise whether associations are worthy of note, estimate the probability that an effect will be replicated in other studies, test the goodness of fit of a logistic model, appraise the bias in a survival curve because of dropouts, and examine the effects of publication bias and the inclusion of specific studies in a meta-analysis. Use is made of WINPEPI in the basic epidemiology and statistics textbooks by Gerstman (2008, 2013) and LaTorre (2010).

Self-instruction

Many students find self-instruction (self-directed learning) attractive because they can choose their own pace and sequencing, working alone or with friends. Some students need only be 'pointed in the right direction' and can then be largely left to their own devices; others require considerable ongoing guidance. An important advantage of self-instruction is that it nurtures the capacity to learn in a self-directed way. Unless students have this skill, they will cease to learn when they complete their formal tuition and are exposed to the 'future shock' they could encounter after graduation . . . 'their world will be characterized by rapid change, and by many facts and skills which have [a] half-life of ten years or less' (Higgs 1982). A major goal of an epidemiology course, it has been said, is 'to provide you with some perspectives and tools that will prevent you from becoming obsolete after you leave medical school' (Ernster 1979).

In theory, self-instruction saves teachers' time. In practice, it may not do so because of the need to prepare appropriate learning guides and other material and to provide supplementary individual tutoring. But if suitable self-teaching aids are already available, the technique may be a valuable one where teachers of epidemiology are scarce or where teachers would like to devote more of their limited time to dialogues with students.

Use may be made of textbooks and journals, selected material obtained from the Web, or full-blown self-instruction packages.

Reading

Self-instruction by reading is, of course, an ancillary method in all epidemiology courses. Students are generally given lists of required and recommended reading. This may be regarded as background learning or it may be given more prominence by requiring its use in class presentations or discussions or in written assignments.

Many epidemiology textbooks are available, and their number is increasing fast. At the last count, Google Books listed 957 books on epidemiology. The books range from basic texts such as *Epidemiology Kept Simple* (Gerstman 2013), whose preface quotes Einstein, 'Things should be made as simple as possible, but not any simpler,' *Epidemiology—An Introduction* (Rothman 2012), *Epidemiology* (Gordis 2008), and *A Pocket Guide to Epidemiology* (Kleinbaum et al. 2006), to the authoritative but challenging *Modern Epidemiology* by Rothman et al. (2008), the imposing *Encyclopedia of Epidemiologic Methods* (Gail and Benichou 2000), and the voluminous *Handbook of Epidemiology* (Ahrens and Pigeot 2005).

Selecting a suitable textbook may not be easy. Many teachers of epidemiology find that there is no single text they can recommend, if only because the objectives of courses vary, as do the contents, emphases, and approaches of textbooks. Some books deal only with particular aspects, for example with basic methods of epidemiological studies (Silman and MacFarlane 2002; Abramson and Abramson 2008), with clinical epidemiology (e.g. Fletcher and Fletcher 2005; Haynes et al. 2006; Weiss 2006; Grobbee and Hoes 2009), social epidemiology (Ibrahim 2004; Cwikel 2006), or field epidemiology (Gregg 2008), or with specific diseases or domains of epidemiology (cited in other chapters of this book). Books that claim to be comprehensive texts generally have gaps, and students are therefore often advised to read specified parts of different textbooks, as well as selected journal articles. Some teachers prepare compilations of chosen materials.

If textbooks are in short supply and library facilities are poor, material available on the Internet may be recommended. For example, the 'Supercourse', a rapidly growing collection of 'lectures'(currently over 5,400) on public health at <http://www.pitt.edu/~super1/> are Microsoft PowerPoint presentations. Similar material is provided by TeachEpi (<http://teachepi.org/>). But it should be noted that non-oral lectures can be expected to have most of the disadvantages and none of the advantages of oral lectures. A randomized trial among medical students has shown, however, that PowerPoint presentations with a supplementary audio feed transmitted knowledge as effectively as live lectures and that

students found them more satisfactory (Spickard et al. 2002). Also, introductory textbooks and other teaching material are available on the Internet, such as the 'Public Health Textbook' available at <http://www.healthknowledge.org.uk/public-health-textbook>. Lists of these resources can be found at (inter alia) <http://www.epidemiolog.net>, <http://www.cdc.gov/excite/library/index.htm>, and other websites listed at <http://blog.discoveryeducation.com/blog/2012/05/06/great-resources-to-teach-epidemiology/>.

Students should of course be encouraged to find material on the Internet themselves. It is arguable that training in literature searching skills should be an integral part of epidemiology courses, especially those on clinical epidemiology and evidence-based medicine or evidence-based public health, although a systematic review has provided only limited evidence that such training is effective (Brettle 2003). In one school, a course (with the participation of librarians) that focused on finding and evaluating the best evidence was found to lead to an increased use of journal literature during the clinical clerkship (Dorsch et al. 2004). It is possibly sufficient to use Google Scholar (Noruzi 2005) for most purposes. However, all students should be introduced, even if only briefly, to ways of conducting an exhaustive search (e.g. for meta-analyses) (Hunt et al. 2000), and should become acquainted with PubMed (http://www.ncbi.nlm.nih.gov/sites/entrez>) and its simplified SLIM interface (<https://pmi.nlm.nih.gov/slim/>); but they should be advised that, unless they have acquired the requisite skills, they should enlist the help of a medical librarian. Google Scholar may be insufficiently complete or up to date to be used for exhaustive searches (Giustini 2005; Burright 2006), although a recent study suggests that it may be good enough to be so used (Jean-Francois et al. 2013).

Self-instruction packages

Self-instruction packages may take the form of a parcel of reading material (ready-made or compiled by the instructor), a set of exercises and other printed aids, a package using audiotapes, videotapes, slides or other components, or a self-instructional computer program. Published 'do-it-yourself' books include study guides (Unwin et al. 1997; Hebel and McCarter 2012) and a self-instruction manual on the interpretation of epidemiological data (Abramson and Abramson 2001). Numerous packages are available, some of them on CD-ROMs or DVDs or on the Internet, and some of them accompanied by printed manuals.

As an example, *Studying Populations*, written by Charles Florey and his colleagues, is a widely used computer-assisted learning package of over eighty exercises on basic epidemiological methods, designed for medical and public health students and practitioners (available at <http://www.dundee.ac.uk/~cdvflore/>). The exercises deal with epidemiological and statistical concepts

and methods, demographic tools, and research methods. ActivEpi, which provides a multimedia learner interactive course on epidemiology (see Kleinbaum et al. 2013); and CDC Train, offered by the Public Health Training Network of the US Centers for Disease Control and Prevention (<http://https://cdc.train.org>). The North Carolina Institute for Public Health offers a periodical, *Focus on Field Epidemiology*, whose issues can be used for computer-based self-study (<http://cphp.sph.unc.edu/focus/index.htm>).

A computer-assisted program should not just be an 'electronic page-turner' that has few advantages over printed material but should exploit the computer's capacity to interact with the user; it should permit the learner to choose topics and their sequence and (where appropriate) to control the level of difficulty, and should carry on a dialogue by providing immediate feedback appropriate to the user's responses (correct or incorrect). Well-designed programs provide specific feedback enabling the student to assess his or her own knowledge and progress and, where necessary, reinforce learning. The program should be user friendly; that is, it should be easy to operate, with lucid and unambiguous text, uncluttered displays, clear and simple instructions, and readily available on-screen help, it should permit the correction of mistakes and the changing of answers, allow for easy transition to previous screens, provide an easy way of quitting the program, and so on (Johnson 1985; Skiba 1985; Kearsley 1989). And, of course, its content should meet the requirements of the epidemiology course.

Increasing use of computer-assisted learning programs is to be expected. For many students, their main attraction is that these programs let them work in a non-threatening private learning environment, in which they can make errors without fear of ridicule from peers or teachers (Larson 1984). In case studies of undergraduate 'electronic' courses (i.e. using Internet material and online statistical programs) on epidemiology and on quantitative methods, feedback from students expressed satisfaction with the experience: 'We learned current and usable skills' and 'It's one thing to be taught something in lectures and even have someone else show it to you, but there's a lot to be gained from when you actually do it. That's where the learning comes in'.

For teachers, the attraction of computer-assisted programs is their cost-effectiveness. Tests of the use of computer-assisted programs in the education of health professionals have shown that students learn as well with these as with more traditional teaching strategies but in one-third to one-half the time (Larson 1984). But teachers who propose to write their own computer programs should be ready to devote considerable time to this. Estimates of the time required to develop one hour of computer-based instruction range from 20 hours (Veloski 1986)—with more time needed for pilot testing, evaluation, and

revision—to 228 hours (Reinhardt 1995) or more. Teachers preparing self-instruction packages should be wary of the unwitting use of copyrighted material (Spallek and Schleyer 1999; Crews 2006).

Problem-oriented projects

Problem-oriented projects pose students with tasks that require the employment of epidemiological knowledge and methods. They provide valuable hands-on experience in addressing real-life or simulated problems. Students may evaluate such projects more favourably than formal teaching—and with no detriment to their examination results (Marantz et al. 1991). In a controlled trial among medical students studying epidemiology, the students using problem-based learning had a stronger grasp of epidemiological principles, were more enthusiastic about epidemiology, and performed equally well in examinations (Dyke et al. 2001). Other studies, too, have reported equivalence in examination results (Khan and Fareed 2001; Krueger et al. 2004). A problem-based course in pharmacoepidemiology for undergraduate science students received high ratings; the students 'relished the course' (Rangachari 2004).

Projects may be assigned to individuals or (more usually) to groups. Group projects have the ancillary advantage of providing experience in teamwork. Most students perform well in group projects, even if there are some who leave the work to their more interested and conscientious colleagues. Whole-class projects may be undertaken, in which tasks are divided among different groups. Ongoing tutoring is usually provided. Especially for graduate students, projects may have an appreciable element of self-instruction; students may be required to crystallize their own objectives and reach their own decisions about methods of investigation. Classroom sessions may be organized to review the planning and progress of the projects, so that learning experiences can be shared. Final reports are generally presented orally to the whole class, as well as in writing. A project may form the basis for a thesis; this is a requirement in some courses, especially those that aim to produce researchers or teachers.

The nature and scope of projects vary, depending on students' and teachers' specific interests, the time available, and other considerations of feasibility. Some projects are restricted to library research, particularly for undergraduate students, who may be asked to find, read, and critically appraise articles with a bearing on a decision required in clinical or community health care. Most projects involve systematic data-gathering, in the field or from clinical records (which may be particularly attractive to medical students). The project may be concerned with the health status of a group or community, the determinants of health, or the provision of health care, providing a true experiential-learning

opportunity, and may be descriptive or analytical; it is often evaluative in nature.

Students are generally asked to make recommendations based on their findings, in order to provide experience in the interpretation and use of epidemiological findings. Students of public health (Gofin et al. 1985) and health management (Mercenier and van Balen 1973) may be required to make practical use of the findings for planning or evaluating health care. The use of findings in the development of student-organized action programmes, as well as the controlled evaluation of a programme originated by a previous group of students, have been found to be strong incentives (Schofield and Muller 1973).

Projects sometimes take the form of research aimed at the acquisition of generalizable new knowledge. This kind of experience is a requirement for some graduate students. It has also been advocated as an important and productive contribution to the teaching of epidemiology to undergraduates, sometimes leading to publications. But even with grant support for an adequate infrastructure of personnel and equipment, several generations of students may have to work on each study, and faculty members may have to complete student-initiated projects (Grufferman et al. 1984). Requiring medical students to write a research protocol has been advocated as a method of teaching research design, even if there is no intention to perform the study (Linskey et al. 1987); projects in research design have also been included in undergraduate nursing courses (Laschinger et al. 1990). At a school in India, first-year clinical students conducted case-control studies, and second-year students undertook research projects in community medicine (Soudarssanane et al. 1994). In West Virginia, a course that focused on the writing of a research proposal was found to be more effective and acceptable to medical students than previous approaches to the teaching of epidemiology and led in some instances to actual research (Garland and Pearson 1989).

Project work is now part of the curriculum in public health in the majority of medical schools in the UK (Gillam and Bagade 2006).

Distance learning

As communication technologies advance, distance learning (characterized by the physical separation between instructors and learners) is growing in popularity, and numerous academic programmes are in existence, provided by the Open University and other institutions. Emphasis has been given to the role of distance learning in epidemiology for students and professionals in countries where teachers are scarce or students have difficulty in reaching teaching centres, and to its potential for enhancing practical and research collaboration between public health workers in developing and developed countries.

Traditional correspondence courses are being replaced by methods that range from the placing of 'lectures' on the Internet, and electronic mail, to more elaborate methods of transmission and mechanisms that allow real-time interaction between teachers and students or among students. Student discussions may be facilitated by setting up mailing lists, by establishing discussion boards on websites, or even by audio/video teleconferencing. The curriculum may include self-instruction components, problem-solving, and other exercises, and even group projects. Distance learning courses may incorporate computer-assisted learning materials (which imply interaction between learner and computer); but a key characteristic of these courses is that there should also be interaction between learner and teachers—and sometimes between learners as well.

The development of one such course is described by Bruce et al. (2002), who emphasize that its language should be simple and friendly, that the students' workload should be kept to the essential minimum, and that the contents should be non-offensive in all communities. Students particularly liked the Web conferencing system, which was 'highly effective in overcoming the potential isolation of distance study'.

The preparation of material for distance learning may be very time-consuming, and teachers have reported that 'providing effective feedback for students required nearly as much time as in an ordinary course with the same number of students' (Ostbye 1989).

Active student interaction may be encouraged by the posing of thought-provoking questions and by requests for feedback. Common problems include slow feedback from instructors, excessive length of the course, and students' reluctance to participate in online discussions.

Teachers interested in distance learning should be prepared to learn new skills or to collaborate with experts on technical aspects. A useful primer on distance learning in public health is offered by the Public Health Training Network of the US Centers for Disease Control and Prevention. It reports that practical experience indicates that the best distance learning programmes combine a variety of media, selected for different purposes, and that success is enhanced if students are organized into groups under a local facilitator, especially if the groups meet periodically. The primer describes concepts and strategies, and provides references to available resources. Among the methods listed are audio conferences, computer conferences, audiographics, and one-way and two-way video.

Combined methods of teaching

The most effective teaching programmes are probably those that use a variety of different approaches, exploiting the special advantages of each. They may

include courses dealing with different topics and using different teaching methods; with good sequencing and coordination, these can reinforce each other.

A type of integrated course that is well accepted by students, and that I most enjoy teaching, is a 'workshop' course offered as an elective (which ensures that its participants are motivated) to public health students who have completed their basic epidemiology course. It is appropriate for a group of preferably not more than twelve or thirteen, but at a pinch up to, say, sixteen, students. The key elements are reading assignments (two to four papers that are handed out to all the students, for reading before the next week's session), group discussions, and computer exercises. The students are asked to bring their laptop computers, with WINPEPI programs installed. The papers include recent ones dealing with important topics of current interest, as well as older ones chosen because of their teaching potential. They include reports on ecological and other descriptive studies, appraisals of screening or diagnostics tests, case-control studies, cohort studies, clinical and community trials, and meta-analyses.

To ensure interest, the week's assignment usually includes at least one bizarre paper, dealing, for example, with the positive correlation between infant mortality and the prevalence of doctors (Cochrane et al. 1978), a meta-analysis of descriptions of the Indian rope trick (Wiseman and Lamont 1996), a purported relationship between antiperspirants and breast cancer (McGrath 2003), an association between heartburn in pregnancy and the baby's hairiness (Costigan et al. 2006), an association between shaving and stroke (Ebrahim et al. 2003), a systematic review of randomized controlled trials of parachute use (Smith and Pell 2003), a study of the validity of a dog's diagnosis of *Clostridium difficile* infection in hospital patients (Bomers et al. 2012), and randomized controlled trials of the effect of accelerated ice-cream eating on headaches (Kaczorowski and Kaczorowski 2002) and of the effect of prayer on the duration of hospital stay of septicaemia patients discharged five or more years before the praying (Leibovici 2001).

The discussions centre on the methods of study, the manner of drawing inferences, and practical implications. Good studies serve as models, flawed ones as object lessons. The discussions are interspersed with or followed by computer exercises, in which the students perform some of the procedures reported or called for by the papers (estimations of sample size or power, significance tests, estimation of confidence intervals, selection of random samples, randomization, allowance for multiple significance testing, computation of attributable fractions, fail-safe *n,* number needed to treat, direct and indirect standardization, Mantel–Haenszel procedures, multiple logistic analysis, etc.).

Since not all the students bring laptops, they usually work in small groups, with the instructor moving from group to group. From time to time they come

together as a single group, focusing on results explained on a whiteboard or projected on a screen. The computer exercises enable the students to understand the statistical procedures better, and give them confidence in their capacity to do them themselves, in their own later epidemiological studies. As Confucius said, 'I hear and I forget; I see and I remember; I do and I understand.' Well-chosen examples may also open their eyes to awkward truths—students never fail to be surprised to find that an observed prevalence of 120 per 1,000, using a measure whose sensitivity and specificity are 90 per cent, points to a true prevalence of only 25 per 1,000, or to see how inaccurate their guesses about the effect of misclassification on an odds or risk ratio can be.

Teaching in the hospital

Teaching epidemiology in the hospital is of special importance for medical students and others whose main concern is the clinical care of patients, and for students with a special interest in hospital services.

For the former category, the challenge is to bring the students to realize the importance of epidemiology by emphasizing epidemiological topics that are closely related to the clinical problems they encounter. In some medical schools, epidemiologists participate in ward rounds and case discussions or run ward rounds themselves, in order to ensure coverage of the epidemiological features of common diseases and the principles of clinical epidemiology or of evidence-based medicine (i.e. the use of epidemiological knowledge and methods in diagnosis, in prognosis, and in decisions about care). This 'bedside' approach may depend for its success on the attitude of key clinicians to epidemiology, and on the availability of teachers who are versed in clinical medicine as well as in epidemiology. The bedside approach is not necessarily effective, even for teaching its limited epidemiological agenda. In a London medical school, where bedside teaching by epidemiologists was replaced by a seminar course based on small-group sessions and practical exercises, the new course was positively evaluated by students and led to a greater appreciation of the importance of epidemiology for clinical practice (Heller and Peach 1984).

Problem-oriented projects for medical students might take the form of 'literature critiques' on problems that interest them, or the performance of surveys based on samples of hospital patients and dealing with clinical problems.

For public health students and others with a special interest in hospital services, the hospital situation offers a wide variety of opportunities for experience in the application of epidemiological methods. Problem-oriented projects may deal with such topics as nosocomial infections, physical hazards in the hospital environment, sickness absences of personnel, factors influencing the use of services

and compliance with care, the effectiveness and safety of various diagnostic and therapeutic services, and the impact of the hospital on community health.

Teaching in the field

Field projects

Field projects can provide valuable learning experiences. The projects may be undertaken by individual students or by small or large groups. They generally involve surveys of diseases or health status in selected samples or population groups, surveys of environmental and other determinants of health, or studies of health care. Use may be made of available data as well as, or instead of, data gathered by the students. Possible topics range over the whole field of public health. For most students, the experience is most interesting if the project has the aim of solving or alleviating a known health problem.

Tutoring is generally a combined responsibility of teachers of epidemiology (and maybe biostatistics or other subjects) and the personnel of the services or agencies with whose help the projects are conducted. The latter act as preceptors, and the feasibility of projects often depends on the availability of suitable preceptors.

Field projects have been used for teaching epidemiology to health workers of different kinds, ranging from public health specialists to practical nurses (Kurtzman and Block 1983) and community health workers (WHO 1981), and also as opportunities for bringing together groups of students of different disciplines (e.g. medical students and health inspector students; Bennett 1981—to assist them in learning to function in teams concerned with common tasks).

Field projects based on local community services can serve useful educational purposes for all students but students with a specific interest in public health services at a regional or national level may require projects that go beyond a local community. 'Learning while doing' may be the central feature of the training of health professionals for careers in applied epidemiology and preventive medicine, as in the Epidemiologic Intelligence Service programme of the US Public Health Service, a programme which provides two years of practical experience in field investigations and epidemiologic research of local, state, national or international importance (Thacker et al. 2001) and similar field epidemiology training programmes in other countries.

Teaching in the community

A community health centre (or a general practice or other primary care service that serves a defined population) offers special opportunities for a wide variety of epidemiological projects of interest to various categories of students

(Hannaford et al. 2005). Possible topics include common diseases and disabilities; familial, social, environmental, and behavioural factors and their effects on health and disease; processes of growth, development and ageing; and various aspects of the care of people living in their own homes.

A health centre or service that practises community-oriented primary care (COPC) can be a particularly useful teaching context. COPC combines the care of individuals and families with the care of the community as a whole, in a single integrated practice, and is characterized by the development of programmes that deal systematically with the health needs of the community and its subgroups (Kark 1981; Abramson 1988; Kark et al. 1994; Rhyne et al. 1998; Gofin and Gofin 2011). It is an evidence-based form of health care, using epidemiology as the basis for the appraisal and elucidation of needs and for the planning and ongoing monitoring and evaluation of the programmes—that is, epidemiology in practice, incarnate. Attractive features for medical students are the emphasis on a real community problem (a 'community case') and the similarity of the COPC cycle (examination and community diagnosis, consideration of different interventions and their probable effectiveness, intervention, surveillance, evaluation, and modification of intervention) to the approach they have learnt to apply to individual patients in their clinical work.

The service is likely to be a useful context for the teaching of epidemiology only if its personnel are interested in the use of epidemiology and if there are ongoing programmes that demonstrate the utility of epidemiology in the planning, provision, or evaluation of care. Epidemiological activities that appear to be conducted solely for teaching purposes may have a negative impact on students.

The opportunities that a COPC practice provides for teaching epidemiology (Kark et al. 1973; Gofin et al. 1985; Boufford and Shonubi 1986; Osborne et al. 1986) may be used in at least four ways. First, students can be exposed to the COPC programmes and thus observe how epidemiology can contribute to community health care. They will see what methods are used—how community diagnosis and health surveillance are conducted, how the findings are used in the planning of community programmes, how the programmes are monitored and evaluated, and how new findings are used as a basis for programme modifications. Second, students can be given 'laboratory' exercises based on data collected in the practice. Third, they can undertake projects related to the centre's activities. These may be directed at collecting information that may help in modifying existing programmes or developing new ones, or they may be less pragmatic 'research'-oriented studies that make use of the centre's accumulated data and other resources. Fourth, students may be asked to apply what they have learnt by planning COPC programmes for the communities in which they themselves

work or will work, using available information about these communities and their health services; problem-solving projects of this kind are best performed by small groups of students, at least one of whom knows the community well.

Integrated workshops on COPC, incorporating most or all of these approaches, together with teaching concerned with basic management skills, community health education, community organization, and other topics, have been provided in Jerusalem for fifty years (Neumark et al. 2011) for public health students, residents in family medicine, paediatricians, and others (Kark et al. 1973; Gofin et al. 1985). In general, they have been favourably received, and most participants subsequently reported that they were applying what they had learnt. In a number of instances, the students' plans for health programmes in other communities were later put into operation.

Field demonstrations

Field demonstrations based on visits to agencies concerned with the collection, analysis, or use of epidemiological data are worthwhile only if they are well prepared and organized, and are clearly concerned with topics of interest to the students. Otherwise, they can easily deteriorate into casual sightseeing trips.

Assessing students' achievements

There are many ways of evaluating students; for example, by using written examinations comprising short or long essay questions, multiple-choice questions, or true/false questions, or oral or take-home examinations, or by marking problem-solving exercises, or by appraising the performance of projects, such as the preparation of a research proposal or a critique of a paper. Different teachers will prefer different methods for courses of different kinds and for students of different kinds.

In the type of small-group 'workshop' course described earlier, where motivated public health students present and discuss published study reports and perform relevant computer exercises, there is a lot to be said for the kind of appraisal commonly used in grading the clinical work of medical students (Magarian and Mazur 1990), namely, a subjective evaluation by the teacher, based here on the student's participation in discussions and other class activities. Experience indicates that when students in such a course are allowed to decide how they will be appraised, they always opt for this method.

Acknowledgements

Text extracts reproduced with permission from *American Journal of Obstetrics and Gynecology*, Volume 191, Issue 2, Krueger, P. M. et al., To the point: reviews

References

Abramson, J. H. (1988) Community-oriented primary care—strategy, approaches and practice: a review. *Public Health Reviews*, **16**: 35–98.

Abramson, J. H. (1997) 'Epidemiology—to be taken with care', in M. Sidell, L. Jones, J. Katz, A. Perbedy, and J. Douglas, eds, *Debates and Dilemmas in Promoting Health*. Houndmills: MacMillan, pp. 143–55.

Abramson, J. H. (2004) WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiologic Perspectives and Innovations*, **1**: 6.

Abramson, J. H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives and Innovations*, **8**: 1.

Abramson, J. H. and Abramson, Z. H. (2001) *Making Sense of Data: A Self-Instruction Manual on the Interpretation of Epidemiological Data* (3rd edn). New York: Oxford University Press.

Abramson, J. H. and Abramson, Z. H. (2008) *Research Methods in Public Health: Surveys, Epidemiological Research, Programme Evaluation, Clinical Trials* (6th edn). Chichester: John Wiley & Sons.

Ahrens, W. and Pigeot, I., eds. (2004) *Handbook of Epidemiology*. Berlin and Heidelberg: Springer.

Bayona, M., Leaverton, P. E., Rangel-Sharpless, M. C., and Williams, P. D. (1994) Short course training in epidemiology and biostatistics for graduate and undergraduate public health professionals. *Public Health Reports*, **109**: 435–8.

Bennett, F. J. (1981) Community diagnosis: its uses in the training of community health workers and in primary health care in East Africa. *Israel Journal of Medical Sciences*, **17**: 129–37.

Billson, J. M. (1986) The college classroom as a small group: some implications for teaching and learning. *Teaching Sociology*, **14**: 143–51.

Bomers, M. K., van Agtmael, M. A., Luik, H., van Veen, M. C., Vandenbroucke-Grauls, C. M., and Smulders, Y. M. (2012) Using a dog's superior olfactory sensitivity to identify *Clostridium difficile* in stools and patients: proof of principle study. *British Medical Journal*, **345**: e7396.

Boufford, J. I. and Shonubi, P. A. (1986) *Community Oriented Primary Care: Training for Urban Practice*. New York: Praeger.

Brettle, A. (2003) Information skills training: a systematic review of the literature. *Health Information and Libraries Journal*, **20 Suppl. 1**: 3–9.

Bruce, J. C., Bond, S. T., and Jones, M. E. (2002) Teaching epidemiology and statistics by distance learning. *Statistics in Medicine*, **21**: 1009–20.

Buck, C., Llopes, A., Nájera, E., and Terris, M., eds. (1988) *The Challenge of Epidemiology—Issues and Selected Readings*. Washington, DC: Pan American Health Organization.

Burright, M. (2006) *Database Reviews and Reports: Google Scholar – Science & Technology*. Available at: <http://www.istl.org/06-winter/databases2.htm1>.

Cochrane, A. I., St Leger, A. S., and Moore, F. J. (1978) Health service 'input' and mortality 'output' in developed countries. *Journal of Epidemiology and Community Health*, **32**: 200–5.

Coomarasamy, A., Taylor, R., and Khan, K. (2003) A systematic review of postgraduate teaching in evidence-based medicine and critical appraisal. *Medical Teacher*, **25**: 77–81.

Costigan, K. A., Sipsma, H. L., and DiPietro, J. A. (2006) Pregnancy folklore revisited: the case of heartburn and hair. *Birth*, **33**: 311–14.

Crews, K. D., ed. (2006) *Copyright Law for Librarians and Educators* (2nd edn). Chicago: American Library Association.

Cwikel, J. G. (2006) *Social Epidemiology: Strategies for Public Health Activism*. New York: Columbia University Press.

Dorsch, J. L., Aiyer, M. K., and Meyer, L. E. (2004) Impact of an evidence-based medicine curriculum on medical students' attitudes and skills. *Journal of the Medical Library Association*, **92**: 397–406.

Dyke, P., Jamrozik, K. P. D., and Plant, A. J. (2001) A randomized trial of a problem-based learning approach for teaching epidemiology. *Academic Medicine*, **76**: 373–9.

Ebrahim, S., Smith, G. D., May, M., and Yarnell, J. (2003) Shaving, coronary heart disease, and stroke: the Caerphilly study. *American Journal of Epidemiology*, **157**: 234–8.

Ernster, V. L. (1979) On the teaching of epidemiology to medical students. *American Journal of Epidemiology*, **109**: 617–18.

Fletcher, R. H. and Fletcher, S. W. (2005) *Clinical Epidemiology: the Essentials* (4th edn). Baltimore, MD: Lippincott Williams & Wilkins.

Ghali, W. A., Saitz, R., Eskew, A. H., Gupta, M., Quan, H., and Hershman, W. Y. (2000) Successful teaching in evidence-based medicine. *Medical Education*, **34**: 18–22.

Gail, M. H. and Bénichou, J. (2000) *Encyclopedia of Epidemiologic Methods* (2nd edn). New York: John Wiley and Sons.

Gange, S. J. (2008) Teaching epidemiologic methods. *Epidemiology*, **19**: 353–356.

Garland, B. K. and Pearson, R. J. C. (1989) Epidemiology course for medical students focuses on proposal writing. *American Journal of Preventive Medicine*, **5**: 240–3.

Gehlbach, S. H., Farrow, S. C., Fowkes, F. G. R., West, R. R., and Roberts, C. J. (1985) Epidemiology for medical students: a controlled trial of three teaching methods. *International Journal of Epidemiology*, **14**: 178–81.

Gerstman, B. B. (2008) *Basic Biostatistics: Statistics for Public Health Practice*. Boston, MA: Jones and Bartlett.

Gerstman, B. B. (2013) *Epidemiology Kept Simple: an Introduction to Traditional and Modern Epidemiology* (2nd edn). Hoboken, NJ: John Wiley & Sons.

Gillam S. and Bagade, A. (2006) Undergraduate public health education in UK medical schools – struggling to deliver. *Medical Education*, **40**: 430–6.

Giustini, D. (2005) How Google is changing medicine. *British Medical Journal*, **331**: 1487–8.

Gofin, J. and Gofin, R. (2011) *Essentials of Global Community Health*. Sudbury, MA: Jones and Bartlett Learning.

Gofin, J., Mainemer, N., and Kark, S. L. (1985) 'Community health in primary care – a workshop on community-oriented primary care', in U. Laaser, R. Senault, and H. Viefhus, eds, *Primary Health Care in the Making*. Berlin: Springer-Verlag, pp. 17–21.

Gordis, L. (2008) *Epidemiology: With STUDENT CONSULT Online Access* (4th edn). Philadelphia, PA: Saunders.

Gregg, M. B. (2008) *Field Epidemiology* (3rd edn). New York: Oxford University Press.

Grobbee, D. E. and Hoes, A.W. (2009) *Clinical Epidemiology: Principles, Methods, and Applications for Clinical Research*. Sudbury, MA: Jones and Bartlett.

Grufferman, S., Kimm, S. Y. S., and Maile, M. C. (1984) Teaching epidemiology in medical schools: a workable model. *American Journal of Epidemiology*, **120**: 203–9.

Hannaford, P. C., Smith, B. H., and Elliott, A. M. (2005) Primary care epidemiology: its scope and purpose. *Family Practice*, **23**: 1–7.

Haynes, R.B., Sackett, D. L., Guyatt, G. H., and Tugwell, P. (2006) *Clinical Epidemiology: How To Do Clinical Practice Research* (3rd edn). Philadelphia, PA: Lippincott Williams & Wilkins.

Hebel, J. R. and McCarter, R. (2012) *Study Guide to Epidemiology and Biostatistics* (7th edn). Sudbury, MA: Jones and Bartlett.

Heller, R. F. and Peach, H. (1984) Evaluation of a new course to teach the principles and clinical applications of epidemiology to medical students. *International Journal of Epidemiology*, **13**: 533–7.

Higgs, J. (1982) 'Self-instructional materials as components of courses', in C. E. Ewan, ed., *Self-Instruction: A Strategy for Education of Health Personnel*. Kensington, New South Wales: University of New South Wales, pp. 2–22.

Hunt, D. L., Jaeschke, R., and McKibbon, K. A. (2000) Users' guides to the medical literature: 21: using electronic health information resources in evidence-based practice. *Journal of the American Medical Association*, **283**: 1875–9.

Hyde, C., Parkes, J., Deeks, J., and Milne, R. (2000) *Systematic Review of Effectiveness of Teaching Critical Appraisal*. Oxford: Institute of Health Sciences, Centre for Statistics in Medicine.

Ibrahim, M. A. (2004) *Social Epidemiology*. Oxford: Oxford University Press.

Ioannidis, J. P. A. (2005) Why most published research findings are false. *PLoS Medicine*, **2**: e124.

James, E. L., Graham, M. L., Snow, P. C., and Ward, B. M. (2006) Teaching research and epidemiology to undergraduate students in the health sciences. *Australian and New Zealand Journal of Public Health*, **30**: 575–8.

Jean-Francois, G., Laetitia, R., and Stefan, D. (2013) Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making*, **13**: 7.

Johnson, A. T. (1985) User friendliness in microcomputer programs. *Computer Programs in Biomedicine*, **19**: 127–30.

Kaczorowski, M. and Kaczorowski, J. (2002) Ice cream evoked headaches (ICE-H) study: randomised trial of accelerated versus cautious ice cream eating regimen. *British Medical Journal*, **325**: 1445–6.

Kark, S. L. (1981) *The Practice of Community-Oriented Primary Care*. New York: Appleton-Century-Crofts.

Kark, S. L., Abramson, J. H., and Gofin, J., eds. (1994) *Atencion primaria orientada a la comunidad (APOC)*. Barcelona: Ediciones Doyma S.A.

Kark, S. L., Mainemer, N., Abramson, J. H., Levav, I., and Kurtzman, C. (1973) Community medicine and primary health care: a field workshop on the use of epidemiology in practice. *International Journal of Epidemiology*, **2**: 419–26.

Kearsley, G. (1989) Good versus bad software: what makes the difference? *Computers in Life Science Education*, **6**: 1–3.

Khan, L. and Fareed, A. (2001) Problem-based learning variant: transition phase for a large institution. *Journal of Pakistani Medical Association*, **51**: 268–70.

Khoury, M. J., Gwinn, M., and Ioannidis, J. P. A. (2010) The emergence of transitional epidemiology: from scientific discovery to population health impact. *American Journal of Epidemiology*, **172**: 517–24.

Khoury, M. J., Millikan, R., Little, J., and Gwinn, M. (2004) The emergence of epidemiology in the genomics age. *International Journal of Epidemiology*, **33**: 936–44.

Kleinbaum, D. G., Sullivan, K. M., and Barker, N. D. (2006) *A Pocket Guide to Epidemiology*. New York: Springer.

Kleinbaum, D. G., Sullivan, K. M., and Barker, N. D. (2013) *ActivEpi Companion Textbook* (2nd edn). New York: Springer.

Kreis, A., Orvad, A., Ruberu, D., and Stace, R. (1998) *Applied Epidemiology – A Full-Subject Self-Directed Computer-Based Problem-Solving Learning Experience*. <http://ro.uow.edu.au/asdpapers/47/>, accessed November 2014.

Krueger, P. M., Neutens, J., Bienstock, J., Cox, S., Erickson, S., Goepfert, A., Maya Hammoud, M., Hartmann, D., Puscheck, E., and Metheny, W. (2004) To the point: reviews in medical education teaching techniques. *American Journal of Obstetrics and Gynecology*, **191**: 408–11.

Kurtzman, C. and Block, D. (1983) Preparation of nurses for community orientation in primary health care in Israel. *Israel Journal of Medical Sciences*, **19**: 768–70.

Larson, D. E. (1984) Using computer-assisted instruction in the education of health care professionals: what the dean needs to know. *Computers in Life Science Education*, **1**: 65–7.

Laschinger, H. S., Johnson, G., and Kohr, R. (1990) Building undergraduate nursing students' knowledge of the research process in nursing. *Journal of Nursing Education*, **29**: 114–17.

La Torre, G. (2010) *Applied Epidemiology and Biostatistics*. Torino: SE*Ed.*

Leibovici, L. (2001) Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. *British Medical Journal*, **323**: 1450–1.

Linskey, M. E., Neugut, A. I., Hall, E., and Cox, J. D. (1987) A course in medical research study design and analysis. *Journal of Medical Education*, **62**: 143–5.

Lowe, C. R. and Kostrzewski, J., eds. (1973) *Epidemiology: A Guide to Teaching Methods*. Edinburgh: Churchill Livingstone.

Magarian, G. J. and Mazur, D. J. (1990) Evaluation of students in medical clerkships. *Academic Medicine*, **65**: 341–5.

Marantz P. R., Burton, W., and Steiner-Grossman, P. (2003) Using the case-discussion method to teach epidemiology and biostatistics. *Academic Medicine*, **78**: 365–71.

Marantz, P. R., Croen, L., Wassertheil-Smoller, S., and Lukashok, H. (1991) Teaching clinical epidemiology to medical students using a collaborative learning model. *American Journal of Preventive Medicine*, **7**: 121–3.

McGrath, K. G. (2003) An earlier age of breast cancer diagnosis related to more frequent use of antiperspirants/deodorants and underarm shaving *European Journal of Cancer Prevention*, **12**: 479–85.

Mercenier, P. and van Balen, H. (1973) The experience, development and evaluation of the International Course in Health Management. *International Journal of Epidemiology*, **2**: 129–35.

Moffat, M., Sinclair, H. K., Cleland, J. A., Smith, W. C. S., and Taylor, R. J. (2004) Epidemiology teaching: student and tutor perceptions. *Medical Teacher*, **26**: 691–5.

Neumark, Y., Manor, O., and Berry, E. M. (2011) Promoting public health workforce training for developing and transitional countries: fifty-year experience of the Braun School of Public Health and Community Medicine, Jerusalem, Israel. *Public Health Reviews*, **33**: 251–263.

Noah, N. D. (2001) 'Exercises', in J. Olsen, R. Saracci, **and** D. Trichopoulos, eds, *Teaching Epidemiology: A Guide for Teachers in Epidemiology, Public Health and Clinical Medicine* (2nd edn). Oxford: Oxford University Press, pp. 359–71.

Norman, G. R. and Shanon, S. I. (1998) Effectiveness of instruction in clinical appraisal (evidence-based medicine) skills: a critical appraisal. *Canadian Medical Association Journal*, **158**: 177–81.

Noruzi, A. (2005) Google Scholar: the new generation of citation indexes. *Libri*, **55**: 170–80.

Öcek, Z. A., ÇiÇeklioglu, M., Gürsoy, S. T., Aksu, F., Soyer, M. T., Hassoy, H., Ergin, I., Sayiner, A., and Kandiloglu, G. (2008) Public health education in Ege University Medical Faculty: developing a community-oriented model. *Medical Teacher*, **30**: e180–8.

Ogino, S., King, E. E., Beck, A. H., Sherman, M. E., Milner, D. A., and Giovannucci, E. (2012) Interdisciplinary education to integrate pathology and epidemiology: towards molecular and population-level health science. *American Journal of Epidemiology*, **176**: 659–67.

Osborne, E. H. S., Hearst, N., Lashof, J. C., and Smith, W. M. (1986) Teaching community-oriented primary care (COPC): a practical approach. *Journal of Community Health*, **11**: 165–71.

Ostbye, T. (1989) An 'electronic' extramural course in epidemiology and medical statistics. *International Journal of Epidemiology*, **18**: 275–9.

Rangachari, P. K. (2004) Exploring the context of drug use: a problem-based learning course in pharmacoepidemiology for undergraduate science students. *Naunyn-Schmiedeberg's Archives of Pharmacology*, **369**: 184–91.

Reinhardt, A. (1995) New ways to learn. *Byte Magazine*, **20**: 50.

Rhyne, R., Bogue, R. J., Kukulka, G., and Fulmer, H. (1998) *Community-Oriented Primary Care: Health Care for the 21st Century*. Washington, DC: American Public Health Association.

Riegelman, R. K., Albertine, S., Persily, N. A., Kaelin, M. W., and Cashman, S. (2010) *Curriculum Guide for Undergraduate Public Health Education, Version 3.0*. <http://www.aptrweb.org/resource/resmgr/undergraduateph/curriculum_guide_version3.pdf>, accessed 29 October 2014.

Rothman, K. J. (2012) *Epidemiology—An Introduction*. New York: Oxford University Press.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008) *Modern Epidemiology* (3rd edn). New Delhi: Wolters Kluwer/Lippincott Williams & Wilkins.

Sackett, D. L., Richardson, W. S., Glasziou, P., and Haynes, R. B. (2000) *Evidence-Based Medicine: How to Practice and Teach EBM* (2nd edn). New York: Churchill Livingstone.

Schofield, F. D. and Muller, A. S. (1973) Epidemiology in the undergraduate curriculum of an African medical school. *International Journal of Epidemiology*, **2**: 407–13.

Silman, A. J. and MacFarlane, G. J. (2002) *Epidemiological Studies: A Practical Guide* (2nd edn). Cambridge: Cambridge University Press.

Simpson, J. M. (1995) Teaching statistics to non-specialists. *Statistics in Medicine*, **14**: 199–208.

Skiba, D. J. (1985) Evaluation of computer-assisted instruction courseware. *Computers in Life Science Education*, **2**: 11–14.

Smith, G. C. S. and Pell, J. P. (2003) Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *British Medical Journal*, **327**: 1459–61.

Soudarssanane, M. B., Rotti, S. B., Roy, G., and Srinivasa, D. K. (1994) Research as a tool for the teaching of epidemiology. *World Health Forum*, **15**: 48–50.

Spallek, H. and Schleyer, T. K. L. (1999) Educational implications for copyright in a digital world. *Education*, **63**: 673–81.

Spickard, A., Alrajeh, N., Cordray, D., and Gigante, J. (2002) Learning about screening using an online or live lecture: does it matter? *Journal of General Internal Medicine*, **17**: 540–5.

Taylor, R., Reeves, B., Ewings, P., Binns, S., Keast, J., and Mears, R. (2000) A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Medical Education*, **34**: 120–6.

Thacker, S. B., Danenberg, A. L., and Hamilton, D. H. (2001) Epidemic Intelligence Service of the Centers for Disease Control and Prevention: 50 years of training and service in applied epidemiology. *American Journal of Epidemiology*, **154**: 985–92.

Unwin, N., Carr, S., Leeson, J., and Pless-Mulloli, T. (1997) *An Introductory Study Guide to Public Health and Epidemiology*. Buckingham: Open University Press.

Veloski, J. J. (1986) 'The integration of the computer into medical education', in J. Javitt, ed., *Computers in Medicine: Applications and Possibilities*. Philadelphia, PA: W.B. Saunders Co., pp. 134–53.

Weiss, N. (2006) *Clinical Epidemiology: The Study of the Outcome of Illness* (2nd edn). New York: Oxford University Press.

WHO. (1981) *The Use of Epidemiology by Front-Line Health Workers in Developing Countries*. Unpublished WHO document SHS/SPM/81.3.

Williamson, E., Morley, R., Lucas, A., and Carpenter, J. (2011) Propensity scores: from naive enthusiasm to intuitive understanding. *Statistical Methods in Medical Research*, **21**: 273–93.

Wiseman, R. and Lamont, P. (1996) Unravelling the Indian rope-trick. *Nature*, **383**: 212–13.

Part 4

Pedagogies

Chapter 28

Guide for teaching assistants in a methods course at a department of epidemiology

Naomi Greene and Tarun Bhatnagar

Introduction to the guide for teaching assistants in a methods course in an epidemiology department

The core epidemiologic concepts that are the subject of any methods course may appear clear-cut and simple on the surface but are, in fact, multilayered with many finely distinguished subtleties. It is imperative that students new to the field be articulately walked through the material and not be misinformed. Having a firm command of these concepts is a vital prerequisite for becoming a teaching assistant (TA) for any methods course. The duties and responsibilities of a TA vary according to the educational institution, level of study (i.e. graduate vs undergraduate), and area of concentration (i.e. for majors or non-majors) but there is a common thread that runs true regardless of the situation: the TA represents the bridge between the professor and the student. The TA has the advantage of having *both* a solid understanding of the concepts being presented and having the recent memory of what it was like to hear about and be expected to internalize those concepts for the very first time. This unique position enables the TA to find ways to bring closer together the experience and expertise of the professor and the open and eager minds of the students in ways that allow information to be appreciated and understood.

Especially in the case of a methods course with a very large number of students, the TA provides 'the human contact and personal motivation that can make the difference between success and failure for individual students' (Streichler 2005). However, being a TA does not mean that one has to have all the answers. The TA that can respond to a student's question with 'I am not entirely sure of the answer but I will research this and come back to you with an answer' is one that will earn the trust and respect of his or her students. It can be encouraging and motivating for students beginning to learn complex concepts to work

with and be taught by someone who has had to struggle a bit to gain understanding. Students who feel their TA is working alongside them, rather than speaking down to them, may feel freer to ask questions, volunteer answers that may be wrong, or ask that comments be repeated or rephrased.

Roles and responsibilities

People considering becoming TAs are generally themselves students, still working toward an advanced degree. Along with the demands of being a graduate student, becoming a TA is a substantial responsibility. Taking on the duties of a TA may limit the number of courses a graduate student can reasonably undertake simultaneously, perhaps delaying slightly, completion of coursework. Nevertheless, there are benefits to being a TA that may outweigh the temporary loss of momentum in one's own education pursuits.

First, there is no better way to uncover areas in which one may not have as solid an understanding as previously thought, than to teach someone else. In preparing lesson plans for TA-led discussion sections, this shaky ground soon becomes apparent to the TA, who must work through the particular concept himself or herself or ask for clarification from the professor. Then the TA can teach students with a fresh and confident perspective. Second, being a TA allows a person to experience teaching and can afford an opportunity to decide whether a career in education may be something s/he would want to pursue down the road. Third, working as a TA also allows one to demonstrate the ability to make independent decisions, adapt one's teaching style to the students' needs, and be an effective team player. If successfully negotiated, this can go a long way in establishing one's credibility to work at various positions in a professional career across areas of education, research, or services.

The variety of instructional duties of TAs may include providing logistical support to the professor in charge of the course, attending lectures conducted by the professor, conducting discussion sections that supplement the lectures, assisting with the preparation, proctoring, and grading of examinations, evaluation of homework assignments and student projects, and holding office hours. We will elaborate on each of these and other roles of a TA in this chapter.

When considering taking on the challenge of teaching assistantship, one must be willing and fully prepared to devote approximately twenty hours per week to address adequately all the aspects of the position. This figure could vary with the number of weekly lectures and discussion sections but is a good estimate of the average amount of time required for the combination of lesson planning, attendance of lectures, office hours, email correspondence, meetings with the professor, and post-section analysis.

Fulfilling these multiple roles effectively will require efficient time management on the part of the student-TA. This would involve learning to say 'No' to tasks outside reasonable expectations from a TA, prioritizing the week's activities, making a to-do list for the day, and not ignoring other personal activities such as research, recreation, and socialization. It is always helpful to consult other experienced TAs for ideas and suggestions about teaching and juggling multiple roles.

Before instruction begins

The TA–professor consultation

The most important event to take place before the first day of class is one or more meetings between the professor and the TA. For the TA to be a successful assistant to both the professor and the class, it is vital that s/he understands clearly the depth and breadth of the core methods that the professor plans to cover. A list of covered topics may be available from previous years that can be used as a jumping-off point, but the TA must be able to visualize the progression and pace of how the current course will unfold. This will aid the TA in preparing the course syllabus as well as in working out a skeleton plan for the discussion sections s/he will be leading. In addition, the TA may be able to offer the professor constructive feedback from when s/he was a student, in terms of a concept that was particularly difficult to grasp or perhaps an alternate reading or other resource that the TA found to be helpful when s/he was the student.

Second, the TA will want to know what the professor's expectations are for the discussion sections and, in general, for the role the TA will play. Some professors may wish to suggest topics for discussion or exercises to be worked through, while others may wish to leave the TA to decide on the structure and content of the discussion sections. In either case, before instruction begins, the TA should have a firm idea of what to expect from the professor and what is expected of him or her.

Finally, the method for developing and grading exercises and exams should be outlined before the first day of class (i.e. who will write the homework exercises and exams, when will they occur, who will grade them, who should be in a position to listen to student questions about the grading, and who should be responsible for final decisions on grading). This will allow a clear representation of authority to be established for all involved, the professor, the TA, and the class.

The course syllabus

Creation of the course syllabus, generally undertaken in conjunction with the professor, is the next step. The syllabus should be a fairly uncluttered document that contains information to help students quickly find where and how to contact

the professor and the TA, when and where the class takes place, when and where office hours are held, what topics are planned for each day of class, what are the required readings for each class, due dates for homework exercises or papers, and when and where exams will be held. The TA prepares a draft of the course syllabus and submits this to the professor who should have a chance to review and make changes to content, format, and layout. Once the final version of the course syllabus is agreed upon, it should be printed out in sufficient numbers such that each student can be given a copy, along with a copy for the professor, the TA, the administrative office, and five to ten extras for last-minute enrolees. Any additions or changes to the syllabus during the course should not simply be verbally announced to the students. It is always better to give the students a revised printed syllabus or a printed addendum. If available, use of a web-based course management system (CMS) allows updates to be made and posted instantly with potentially automated notification to users (i.e. students, the professor, and the TA, in this case) that changes have been made.

Online teaching aids

A class website can be a helpful tool for both teachers and students and there are a variety of ready-made CMS available. If the educational institution does not subscribe to any of the available CMS but provides access to a server, a class website can be constructed that can link from the professor's homepage or directly from the department's list of offered courses on the official website. The website should be password protected so that only enrolled students are given permission to access the course materials and readings. If it is password protected, published articles from journals can be made available for downloading and/or printing in PDF (Portable Document Format) format without violating copyright issues. In addition, with special permission, the professor's lecture slides, as yet unpublished documents, and personal communications can be shared with students who are enrolled in the class. Each professor and/or guest lecturer should be given the option of having his or her slides posted before the lecture, or postponing until after. The professor's decision is made based on concerns that students will simply download the slides and not attend the lecture. Within the course syllabus, the title of each required reading can be linked to the PDF file on the university's server so that students may view or print off the required course readings. This saves time in terms of each student having to search individually for articles by name, with success at locating the article dependent on the student's computer skills and access to high-speed Internet. It also saves the department money, time, and resources by avoiding the necessity to needlessly print course materials for each student, especially in a class with a large enrolment, typical of many methods courses.

Furthermore, this type of online course syllabus allows additional readings to be posted and made available instantly, changes can be made to the lecture schedule to accommodate the pace of each class, last-minute guest speakers can be added, exam dates changed, etc. A class email list should be constructed and maintained by the TA. Notification of changes to required readings, additions and/or corrections to the lecture or exam schedule can thus be easily emailed to all enrolled students. Even so, students should be told on the first day of class to check the online course syllabus frequently to make sure they are kept abreast of changes.

The homework assignments

If the TA will be independently responsible for creating homework assignments, this should be undertaken after discussion with the professor and armed with knowledge of the concepts to be covered in lectures. Homework exercises can be developed that are flexible enough to be adapted to match the complexity level at which each concept is taught. The level of complexity at which each concept is taught may not become apparent until the day of the lecture, as this will depend on the ability of students to digest each concept. For example, when teaching the concept of confounding, the professor may choose, based on his or her feeling about how well the students are grasping confounding on a conceptual level, to forego discussion of calculations of standardized association measure estimates and variance estimators until a later time or perhaps permanently. The TA can prepare a homework exercise regarding confounding in which questions can be inserted or exchanged to reflect the level of complexity covered in class lectures. The following example (Fig. 28.1) is taken from a core methods course in the University of California, Los Angeles (UCLA) School of Public Health's Epidemiology Department, adapted and modified from course notes and examples from Epidemiology 601 (University of Michigan, Prof. Hal Morgenstern) and reproduced here with the kind permission of Prof. Morgenstern. The data is not from a real study but was created by Prof. Morgenstern for teaching purposes. The first set (A) of questions reflects the inclusion of class lectures on standardization and variance estimators, while the second set (B) of questions is drawn from a year in which only a more qualitative approach was taken.

If homework assignments are to be posted on a class website, they should be posted in a rich text format (RTF) so that students can download them, fill in the answers, and print the final version to submit. Once the due date for submitting homework has passed, the homework assignment with suggested answers should be posted to the class website in a read-only format.

A library of homework assignments may be assembled and augmented from year to year. Such an accumulation of homework assignments allows for

A study was done to estimate the acute effect of Type A behavior on the occurrence of premature ventricular beats (PVB) among healthy young adults exposed to a stressful situation. The investigator selected 600 university students, ages 18–25, and categorized them as either Type A or Type B, using a structured interview. After the interview, subjects were given a brief physical exam, including an EKG, and they were asked to solve a difficult puzzle in 30 minutes. The table below shows the number of subjects who experienced PVB (detected by EKG) while trying to solve the puzzle, by behavior type, race and sex. Assume that race and sex are risk factors for PVB.

		Type A		Type B		
Race	Sex	No. with PVB	No. of subjects	No. with PVB	No. of subjects	Total
White	Male	81	90	9	10	100
	Female	1	10	4	40	50
Black	Male	8	60	12	90	150
	Female	68	240	17	60	300
Total		158	400	42	200	600

QUESTION SET A:

a. Estimate the crude risk ratio for PVB, comparing Type A subjects with Typa B subjects.

b. Estimate the four race-sex-specific risk ratios.

c. Estimate the (internally) standardized risk ratio (sRR), controlling for race (ignoring sex), using the Type A group as the standard population. Estimate the 95 percent confidence limits. Does race appear to be a confounder?

d. Estimate the sRR, controlling for sex (ignoring race), using the Type A group as the standard population. Estimate the 95 percent confidence limits for both point estimates. Does sex appear to be a confounder?

e. Estimate the sRR, controlling for both race and sex, using the Type a group as the standard population. Do race and sex appear to be confounders? Explain.

f. How would the choice of another standard population (e.g., the Type B group) change the result of question e?

QUESTION SET B:

a. Estimate the crude risk ratio for PVB, comparing Type A subjects with Typa B subjects.

b. Estimate the four race-sex-specific risk ratios.

c. Stratify on race ignoring sex. Does race appear to be associated with exposure (personality type) in this study? Explain your answer briefly.

d. Stratify on sex ignoring race. Does sex appear to be associated with disease (PVB) in the unexposed (Type B)? Explain your answer briefly.

e. Draw a directed acyclic graph (DAG) for the Race-Exposure-Disease situation in this study based on your answer from part c.

Fig. 28.1 Example taken from a core methods course from the UCLA School of Public Health Epidemiology Department.

Adapted and modified from course notes and examples from Epidemiology 601 (University of Michigan, Professor Hal Morgenstern) and reproduced here with the kind permission of Professor Morgenstern, USA.

additional flexibility, may save time and effort (i.e. avoid 'reinventing the wheel'), and prove a valuable asset and training tool for TAs with less teaching experience. We list references below which offers exercises and suggested answers for many important epidemiologic concepts (Weiss 2012; Johnson and Howards 2013).

Logistic support

Shortly before class begins, certain logistic tasks should be completed in order to ensure a smooth period of instruction. The TA may be required to make photocopies of required material to be handed out and will need to ask the departmental administrative personnel for these procedures. The TA should make sure of the location of the assigned classroom and familiarize himself/herself with the setup of the room. If the classroom is located in an unusual place, it may be helpful to post a directional map on the class website or email one to the enrolled students. Check for the appropriate teaching equipment, such as chalk for chalkboard, pens for whiteboard, audiovisual equipment, microphone, etc. Given the facility and university policy, it is also important to ensure availability of assigned books and reading material at the university bookstore and/or in the library.

Once instruction begins

The lecture session

The TA should plan on attending every lecture session for a variety of reasons, not only to provide the professor with technical and logistic assistance but also to take part in class discussions and ask questions that students may not have asked in order to clarify complicated concepts. In addition, there is no better way for the TA to get a sense of the pace of the class, the breadth and depth of understanding the class demonstrates, than by direct and consistent observation. This will aid the TA in planning meaningful discussion sections and in making unanticipated changes to set lesson plans based on what s/he observes on a given day in the lecture session.

Furthermore, as discussed above, the TA has the advantage of being able to hear the professor's lecture with the student's ear and yet process what is taught with the aid of the greater understanding s/he has developed since being a student. For example, in a lecture that addresses 'confounding' and 'association measure modification', the TA may recognize an ensuing silence on the part of the students as an utter inability to formulate appropriate questions. The TA can raise his or her own hand and pose the question to allow the professor to reiterate or rephrase how these two basic concepts may sound similar at first and yet

how different they truly are. The value of clearly voicing a question that students may want to ask, for which they cannot find the words, cannot be overstated. Setting the example of asking questions in lecture is another way that TAs can earn the respect and trust of their students and facilitate a fruitful class discussion. Some professors may dislike the idea of being questioned by the TA, especially if this questioning is critical. Other professors may welcome an exchange of different points of view because opinions differ in real life.

Taking good notes during lectures is an essential skill for students to learn but in a methods course this can sometimes interfere and detract from the student's ability to fully process and digest core concepts. This can be especially true if the course is taught in a language that is not the student's first language. The TA may wish to obtain a digital recorder and make a recording of each lecture. The digital file can be stored in a variety of formats that can then be easily imported into the wide variety of audio program on computers or personal audio players. The digital sound files can be posted on the class website, each lecture being linked to its sound file that the TA can upload to the server housing the class website. This allows the students to listen to each lecture over and over, in short sections. This frees them to take more skeletally structured notes during class time (along with taking down any figures or tables presented in lecture) and fill in anything they may have missed or been unclear about, by listening to the lecture at their leisure. A more extreme application includes transcribing the lecture and posting the complete lecture notes on the class website, after the professor has reviewed and approved them. This is naturally quite time-consuming, both for the TA, who transcribes the notes, and for the professor, who must review and make any desired changes. However, especially in a situation in which a significant proportion of students are not native speakers, this is an option that may allow all students a more equal footing in completing assignments and passing exams.

Students should be given the opportunity to study or reflect ahead of time on the topic to be presented in the next lecture, and thus the lecture slides, along with suggested resources for further edification, can be posted on the class website in advance of each lecture. Whereas some may be concerned that posting the slides ahead of time will encourage students to miss the lecture, feeling they got what they needed from reviewing the slides, at the university level students should already have discovered that one's education is only as good as the effort one puts into it. The added benefit to conscientious students outweighs the liability to the few whose priorities are ill-advised.

In an effort to make the learning process an interactive one, the TA can develop ways to force students to reflect on each lecture and reading. One technique, which can be called the 'question of the day', was developed by a conscientious

and creative TA at the UCLA School of Public Health, Dr Amy Kaji, and is reproduced here with her permission (personal communication). Essentially, the TA reviews the lecture notes and required readings and formulates one question each day which is emailed to the students, along with the suggested answer. This encourages active learning instead of passive attention to lectures and readings, as the students can test themselves before reading the answer and bring questions to the discussion section. This helps to uncover subject areas that students may not have understood but do not ask about during the lectures and may also provide a direction to the proceedings of the discussion sections. The following are some examples of such questions:

Question 1: Regarding directed acyclic graphs (DAG), which of the following is false?

(a) All of the edges in a DAG are directed (are arrows), and a DAG is acyclic (there are no directed paths that form a closed loop).

(b) In a causal DAG, a marginal association between an exposure and an outcome has only two (possibly both) sources: (1) the exposure is a cause of the outcome (there is a directed path between exposure and disease), and (2) the exposure and outcome share a common cause (there is a backdoor path from exposure to disease).

(c) In a causal diagram, an absence of a directed edge from a variable, A, to another variable, B, represents the assumption that A is not a direct cause of B.

(d) Two variables in a causal DAG are unassociated if there is no open path between them (there is no causal or backdoor path).

(e) If there is an open path between variables A and B, then A and B must be marginally dependent.

Answer: (e) If there is no open path between two variables A and B, then the two variables are marginally independent. Conversely, if the two variables are marginally associated, there must be an open path between them. However, the presence of an open path between A and B does not necessarily mean A and B are marginally dependent, since the associations along different paths can be in opposite directions and 'cancel out' (thus leaving no marginal association). Thus, the presence of an open path does not imply a marginal association between two variables.

Question 2: Regarding confounding, which of the following statements is true?

(a) Confounding is bias in estimation of the exposure effect, due to a lack of comparability between exposed and unexposed groups in the source population.

(b) A necessary property of a confounder is that it must be a risk factor for the disease (or a proxy for the risk factor) in the unexposed source population.

(c) A necessary property of a confounder is that it must be associated with exposure status in the total source population, and its association with the exposure cannot be due entirely to the effect of the exposure on the covariate (an intermediate).

(d) The primary basis for identifying confounders in a given study is prior knowledge of relevant effects or associations in the source population; one cannot merely rely upon the statistical associations observed in the data.

(e) All of the above.

Answer: (e) All of the above statements are true.

Meetings with the professor

In addition to the logistical, administrative, and didactic support that the TA provides the professor, the TA should also serve as a messenger or ombudsman between the students and the professor. Weekly or monthly meetings between the professor and TA(s) provide a good forum for discussing the progress the class is making from both perspectives. The TA should feel welcome in bringing back to the professor the frustrations and concerns that were stated by the students or perceived by the TA. It can only be to the benefit of all concerned to have a frank discussion of what worked and what did not work in the lecture sessions. Whereas many universities have a student evaluation form that is submitted at the end of the term, a dynamic process of updating the professor during the term allows for a greater chance of student success and satisfaction.

The discussion section

A large class should be divided into smaller sections of fifteen to twenty students to enable an active one-on-one interaction between the students and the TA:

> Good Discussion Sections provide students with an opportunity to formulate principles in their own words and to suggest applications of these principles. They also help students become aware of, and define problems implied in, lectures or readings (University of California, Los Angeles 2006).

Many universities publish a general guidebook for TAs that outlines and summarizes techniques for engaging students in discussion and interactive participation. Some of these guidebooks are available to download from the Internet, even for TAs in smaller educational institutions which may not offer a guidebook specific to their institution. Such a guidebook can be a helpful aid to TAs with less teaching experience.

Developing a lesson plan for each discussion section is an important tool to help the TA stay organized and on task and provides a useful way to evaluate the success of each section session. While it may be tempting to plan all scheduled discussion sections before instruction begins, this does not allow for necessary changes to the curriculum or level of complexity that the professor decides is best, as the class evolves and reveals its needs. Planning one section at a time based on observation of the class will probably prove to be the best use of the TA's time and resources. It may be useful to lag the topic for each discussion section by one lecture (i.e. use Tuesday's lecture topic as a basis for Thursday's discussion section) to give students time to absorb the material they just heard while giving them an opportunity to solidify their understanding of concepts they learnt about a few days earlier, having had a chance to review and process. There is no set format for a lesson plan, though the TA should always write out the goals for each section to help him or her concentrate on one or two key ideas and to aid in evaluating how well these goals were met.

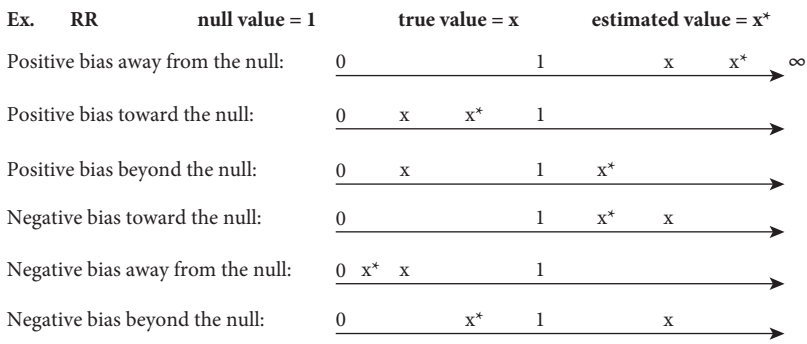
For a core epidemiology methods course, the lesson plan may further include a handout for students summarizing a topic covered in class, with key concepts underlined, assumptions underlying each element in italics, diagrams, or examples from the literature to illustrate, and a list of references to allow students the option of getting multiple perspectives. Figure 28.2 is an example of a summary sheet, including a short exercise used by us at the UCLA School of Public Health.

A handout may instead include an example of a study from the literature, with a series of questions posed (with or without calculations) that the TA and students can expand upon and/or work through during the discussion section. It has been our experience that students appreciate the opportunity to work through multiple concrete examples, whether qualitatively or quantitatively, during the discussion section. The following is an example of a concrete qualitative case: after learning about confounding, students may accept the notion that a single covariate can be thought of as a confounder and an intermediate but will often ask for a 'real-life' example. The TA may respond with the example of a study of the association between physical activity and new onset of wheezing. It may be that, in a particular study, increased physical activity was negatively associated with new wheezing but that the association was reduced or disappeared when controlling for smoking. In one scenario, smoking can be seen as a classic confounder, and in another, one could posit that smoking is mediated by physical activity, on the path from physical activity to new episodes of wheezing. Alternatively, the following is an illustration of a quantitative problem: after learning about standardization in a lecture that included an example of working out an internally standardized risk ratio for a particular

BIAS

Bias is the lack of **internal validity** or incorrect assessment of the association between an exposure and an effect in the target population.
External validity conveys the meaning of generalization of the results observed in one population to others.
There is **no external validity without internal validity**, but the presence of the second does not guarantee the first.
Bias or **systematic error** should be distinguished from **random error** or lack of precision.
Biases can be classified by the direction of the change they produce in an effect parameter.

- away from, toward, or beyond the null value of the effect parameter
- positive or negative, relative to the value of the effect parameter



Specify the direction of bias in each of these studies:

a. The true risk ratio of the relation between an exposure and disease is **2.5** different studies yielded the following estimate of true risk ratio (ignoring random error):

   i) 4
   ii) 1.5
   iii) 0.8

b. True risk ratio is **0.5** and estimated risk ratio is:

   i) 0.2
   ii) 0.9
   iii) 3.0

c. True risk difference is **0.4** and estimated risk difference is:

   i) –0.2
   ii) 0.1
   iii) 0.6

Fig. 28.2 Summary handout for the discussion section.

Adapted and modified from course notes and examples from Epidemiology 601 (University of Michigan, Professor Hal Morgenstern) and reproduced here with the kind permission of Professor Morgenstern, USA.

study, the TA may continue, in the discussion section, with having the class work out the externally standardized counterpart. This leads naturally to a discussion concerning what factors go into the choice of the appropriate standard population and the benefits and limitations of each.

Discussion sections are usually not more than an hour long. It is of the essence that this time be intelligently managed. The TA should ensure that only one to two topics are dealt with per section. The amount of time spent in discussing each topic should be prioritized as per course requirements and level of students' comprehension. More time may be spent on responding to such student questions that are helpful in elaborating and clarifying concepts for most students. In the interest of the group, it is better to deal with individual problems during office hours. Also, administrative issues concerning the class are best taken care of in the lecture period. There are always some students who are more vociferous and others who tend to be inhibited. It is imperative to involve all students in the deliberations. Although volunteering answers from the students should be preferred, it may sometimes be critical to solicit individually directed responses in order to break the inertia.

Do a post-section analysis of how things went—this can be in the form of a journal or can be made into measurable constructs in a spreadsheet, or both.

Pre-examination review

Considering the complexity and breadth of most epidemiology core courses, it is beneficial for the students to have a review session before any of the examinations given in the course. This could be done a couple of days before the scheduled examination date, as it not only refreshes the course material covered prior to the examination but also gives the students an idea about the relative importance of various topics from an examination perspective. There is no set format for a review session. It could simply be a question–answer session where the TA and/or the professor respond to queries from the students. The TA could also prepare and present a summary of all topics on which the students are to be tested. During or after the presentation, students can ask for detailed explanations for specific problem areas. The review could further include a handout summarizing the associated course material, highlighting key concepts that would facilitate the students to focus their studies.

Office hours

TAs are required to hold office hours that should be scheduled in consultation with the students and in an accessible and comfortable place so as to maximize the number of students who can attend. Office hours give a chance to the

students to clarify their queries related to the lectures, course materials, and assignments. Besides, because the students may feel more comfortable with the TA than with a faculty member, it also provides an opportunity to the students to discuss their concerns and apprehensions regarding their progress in the course, along with personal problems. This is where the TA can act as an informal adviser, counsellor, and motivator by providing suggestions or guidance to finding other sources of assistance if s/he feels that the student needs more specific advice. Always encourage students to come to office hours. Create a relaxed and pleasant atmosphere, as some students may feel intimidated by the prospect of meeting face-to-face with a faculty member. Listen attentively to the students and respond thoughtfully so as to use this time effectively. Set clear limits on the amount of time spent with an individual student while others may be waiting. Specify in the syllabus if the TA should hold appointments outside of regular office hours. The students always appreciate extra office hours before examinations, as it gives a last-minute opportunity to clarify study areas and concepts that maybe still be troubling them.

Conclusion

In our experience, the students have expressed greater satisfaction from their TAs and opined that they make 'better teachers'. Compared to the full-time faculty, a TA can be more approachable and available, given that s/he is genuinely concerned about student learning and is friendly and enthusiastic about the subject. Respect the students as learners, and never forget that you went through that same process before. Adapt and change your teaching style, keeping in mind the goals of the course and the varying abilities of students. 'Accept the students and realize that your goal is to help them learn, starting from where they are coming from' (University of California, Los Angeles 2006).

Countries with no current use of TAs in teaching should consider instituting the use of TAs. Ambitious students should become TAs. It is the best way to learn the subject matter well and to be known and respected by the faculty.

Bibliography

The following is a short bibliography of books that address developing and improving the pedagogic skills of TAs.

Althen, G. (1988) *Manual for Foreign Teaching Assistants*. Iowa City, IA: University of Iowa.

Andrews, J., ed. (1985) *Strengthening the Teaching Assistant Faculty*. San Francisco, CA: Josey-Bass.

Bender, T. (2003) *Discussion-based Online Teaching to Enhance Student Learning: Theory, Practice, and Assessment*. Sterling, VA: Stylus Publishing.

Center for Teaching Development, University of California, San Diego. (2014) *The College Classroom*.< https://collegeclassroom.ucsd.edu/teaching-tips/>, accessed 5 December 2014.

Curzan, A. and Damour, L. (2000) *First Day to Final Grade: a Graduate Student's Guide to Teaching*. Ann Arbor, MI: University of Michigan Press.

Davis, B. G. (1993) *Tools for Teaching*. San Francisco, CA: Jossey-Bass.

Gullette, M. M. (1982) *The Art and Craft of Teaching*. Cambridge, MA: Harvard-Danforth Centre for Teaching and Learning.

Johnson, C. Y. and Howards, P. P. (2013) Causal pie bingo! *Epidemiology*, **24**: 331.

McKeachie, W. J. and Svinicki, M. (2006). *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers* (12th edn). Boston, MA: Houghton Mifflin Company.

Smith, J., Meyers, C. M., and Burkhalter, A. J. (1992) *Communicate: Strategies for International Teaching Assistants*. Englewood Cliffs, NJ: Regents/Prentice Hall.

Tice, S. L., Jackson, N., Lambert, L. M., and Englot, P., eds. (2005) *University Teaching: a Reference Guide for Graduate Students and Faculty*. Syracuse, NY: Syracuse University Press.

University of California, Los Angeles. (2006) *The TA Handbook 2006–2007*. Los Angeles: Office of Instructional Development, University of California, Los Angeles.

Weiss, N. (2012) *Exercises in Epidemiology: Applying Principles and Methods*. New York: Oxford University Press.

Index