Pharmacoepidemiology

Pharmacoepidemiology

Sixth Edition

Edited by

Brian L. Strom MD, MPH
Chancellor, Rutgers Biomedical & Health Sciences
Executive Vice President for Health Affairs
University Professor
Rutgers, the State University of New Jersey
Newark, NJ, USA

Stephen E. Kimmel MD, MSCE
Professor of Medicine and of Epidemiology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA, USA

Sean Hennessy PharmD, PhD
Professor of Epidemiology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA, USA

WILEY Blackwell

This edition first published 2020 © 2020 by John Wiley & Sons Ltd

Edition History
5e 2012 John Wiley & Sons Ltd.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data
Names: Strom, Brian L., editor. | Kimmel, Stephen E., editor. | Hennessy, Sean, editor.
Title: Pharmacoepidemiology / edited by Brian L. Strom, Stephen E. Kimmel, Sean Hennessy.
Other titles: Pharmacoepidemiology (Strom)
Description: Sixth edition. | Hoboken, NJ : Wiley-Blackwell, 2020. | Includes bibliographical references and index.
Identifiers: LCCN 2019029935 (print) | ISBN 9781119413417 (cloth) | ISBN 9781119413370 (adobe pdf) | ISBN 9781119413424 (epub)
Subjects: MESH: Pharmacoepidemiology–methods
Classification: LCC RM301.5 (print) | LCC RM301.5 (ebook) | NLM QV 771 | DDC 615.7–dc23 LC record available at https://lccn.loc.gov/2019029935
LC ebook record available at https://lccn.loc.gov/2019029936

Cover images: © jokerpro/iStock.com, © crisserbug/iStock.com, © AlexRaths/iStock.com
Cover design by Wiley

Set in 10/12pt Warnock by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

Contributors *x*
Preface *xix*
Acknowledgments *xxiii*

Part I Introduction *1*

1 What Is Pharmacoepidemiology? *3*
Brian L. Strom

2 Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiologic Studies *27*
Jeffrey S. Barrett

3 Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies *44*
Brian L. Strom

4 Sample Size Considerations for Pharmacoepidemiologic Studies *60*
Brian L. Strom

5 When Should One Perform Pharmacoepidemiologic Studies? *71*
Brian L. Strom

Part II The Role of Pharmacoepidemiology in Different Sectors *81*

6 The Role of Pharmacoepidemiology in the Healthcare System and Academia *83*
Joshua J. Gagne and Jerry Avorn

7 The Role of Pharmacoepidemiology in Industry *98*
Nicolle M. Gatto, Rachel E. Sobel, Jamie Geier, Jingping Mo, Andrew Bate, and Robert F. Reynolds

8 **The Role of Pharmacoepidemiology in Regulatory Agencies** *126*
Gerald J. Dal Pan, June Raine, and Shinobu Uzu

9 **Pharmacoepidemiology and the Law** *140*
Aaron S. Kesselheim and Kerstin N. Vokinger

Part III Sources of Data for Pharmacoepidemiology Research *165*

Part IIIa Spontaneous Reporting *167*

10 **Postmarketing Spontaneous Pharmacovigilance Reporting Systems** *169*
Gerald J. Dal Pan, Marie Lindquist, and Kate Gelperin

Part IIIb Electronic Data Systems *203*

11 **Overview of Electronic Databases in Pharmacoepidemiology** *205*
Brian L. Strom

12 **Encounter Databases** *211*
Tobias Gerhard, Yola Moride, Anton Pottegård , and Nicole Pratt

13 **Electronic Health Record Databases** *241*
Daniel B. Horton, Harshvinder Bhullar, Lucy Carty, Francesca Cunningham, Alexis Ogdie, Janet Sultana, and Gianluca Trifirò

14 **Inpatient Databases** *290*
James A. Feinstein, Peter K. Lindenauer, Chris Feudtner, and Brian T. Fisher

Part IIIc Studies Involving *Ad Hoc* Data Collection *305*

15 **Event Monitoring in the UK** *307*
Vicki Osborne and Saad A.W. Shakir

16 **Primary Data Collection for Pharmacoepidemiology** *342*
Nancy A. Dreyer, Ana Filipa Macedo, and Priscilla Velentgas

Part IIId Choosing a Data Source *355*

17 **Choosing among the Available Data Sources for Pharmacoepidemiology Research** *357*
Brian L. Strom

Part IV Selected Applications of Pharmacoepidemiology *373*

18 Studies of Drug Utilization *375*
Björn Wettermark, Vera Vlahović-Palčevski, David Lee, and Ulf Bergman

19 Evaluating and Improving Physician Prescribing *411*
Christine Y. Lu, the late Sumit R. Majumdar, Helene Lipton, and Stephen B. Soumerai

20 Pharmacoepidemiologic Studies of Vaccine Safety *437*
Robert T. Chen, Jason M. Glanz, and Tom T. Shimabukuro

21 Epidemiologic Studies of Medical Devices: Methodologic Considerations for Implantable Devices *496*
Danica Marinac-Dabic, Sharon-Lise Normand, Art Sedrakyan, and Thomas P. Gross

22 Research on the Effects of Medications in Pregnancy and in Children *524*
Daniel B. Horton, Sonia Hernandez-Diaz, Tamar Lasky, and Krista F. Huybrechts

23 Study of Biologics and Biosimilars *561*
Jeffrey R. Curtis and James D. Lewis

24 Risk Management *581*
Claudia Manzo, Emil Cochino, Lubna Merchant, and Giampiero Mazzaglia

25 Distributed Networks of Databases Analyzed Using Common Protocols and/or Common Data Models *617*
Sengwee Toh, Nicole Pratt, Olaf Klungel, Joshua J. Gagne, and Robert W. Platt

26 Comparative Effectiveness Research *639*
Soko Setoguchi and Ian Chi Kei Wong

27 Data Mining and Other Informatics Approaches to Pharmacoepidemiology *675*
Andrew Bate, Gianluca Trifirò, Paul Avillach, and Stephen J.W. Evans

28 Pharmacoepidemiologic Research on Drugs of Abuse *701*
Jana McAninch, Alex Secora, Cynthia Kornegay, and Judy Staffa

Part V Selected Special Methodologic Issues in Pharmacoepidemiology *723*

29 Assessing Causation from Case Reports *725*
Judith K. Jones, Bernard Bégaud, and Elyse Kingery

30 Molecular Pharmacoepidemiology *746*
Christine Y. Lu and Stephen E. Kimmel

31 Bioethical Issues in Pharmacoepidemiologic Research *772*
Laura E. Bothwell, Annika Richterich, and Jeremy Greene

32 The Use of Randomized Controlled Trials in Pharmacoepidemiology *792*
Robert F. Reynolds, Samuel M. Lesko, Nicolle M. Gatto, Tjeerd P. van Staa, and Allen A. Mitchell

33 The Use of Pharmacoepidemiology to Study Beneficial Drug Effects *813*
Brian L. Strom and the late Kenneth L. Melmon

34 Pharmacoeconomics: The Economics of Pharmaceuticals *837*
Kevin A. Schulman

35 Benefit–Risk Assessments of Medical Treatments *867*
Bennett Levitan, Rachael DiSantostefano, and Scott Evans

36 The Use of Metaanalysis in Pharmacoepidemiology *897*
Jesse A. Berlin, Brenda J. Crowe, H. Amy Xia, and Stephen J.W. Evans

37 Validity of Drug and Diagnosis Data in Pharmacoepidemiology *948*
Mary Elizabeth Ritchey, Suzanne L. West, and George Maldonado

38 Studies of Medication Adherence *991*
Julie Lauffenburger, Trisha Acri, and Robert Gross

39 Risk Evaluation and Communication *1010*
Susan J. Blalock, Rebecca Dickinson, and Peter Knapp

40 Methods for Studying the Health Effects of Drug–Drug Interactions *1030*
Sean Hennessy, Charles E. Leonard, Joshua J. Gagne, James H. Flory, Colleen M. Brensinger, and Warren B. Bilker

41 The Pharmacoepidemiology of Medication Errors *1046*
Hanna M. Seidling and David W. Bates

42 Patient Engagement and Patient-Reported Outcomes *1061*
Esi M. Morgan

43 Advanced Approaches to Controlling Confounding in Pharmacoepidemiologic Studies *1078*
Sebastian Schneeweiss and Samy Suissa

Part VI Conclusion *1109*

44 The Future of Pharmacoepidemiology *1111*
Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy

Appendix A Sample Size Tables *1123*

Appendix B Glossary *1141*

Index *1155*

Contributors

Trisha Acri
Philadelphia, PA
USA

Paul Avillach
Department of Biomedical Informatics
Harvard Medical School
Boston, MA
USA

Jerry Avorn
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
USA

Jeffrey S. Barrett
Bill and Melinda Gates Medical Research Institute
Cambridge, MA
USA

Andrew Bate
Pfizer Ltd
Walton-on-the-Hill, Surrey
UK

David W. Bates
Division of General Internal Medicine and Primary Care
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
USA

Bernard Bégaud
School of Medicine
University of Bordeaux
Bordeaux
France

Ulf Bergman
Division of Clinical Pharmacology
Department of Laboratory Medicine
Centre for Pharmacoepidemiology
Karolinska Institute
Karolinska University Hospital-Huddinge
Stockholm
Sweden

Jesse A. Berlin
Johnson & Johnson
Titusville, NJ
USA

Harshvinder Bhullar
Real-World Insights, IQVIA
London
UK

Warren B. Bilker
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Susan J. Blalock
Division of Pharmaceutical Outcomes and Policy
Eshelman School of Pharmacy
University of North Carolina at Chapel Hill
Chapel Hill, NC
USA

Laura E. Bothwell
Health Sciences Department
Worcester State University
Worcester, MA
USA

Colleen M. Brensinger
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Lucy Carty
Clinical Practice Research Datalink
Medicines and Healthcare Products Regulatory Agency
London
UK

Robert T. Chen
Brighton Collaboration
Task Force for Global Health
Decatur, GA
USA

Emil Cochino
European Medicines Agency
Amsterdam
The Netherlands

Brenda J. Crowe
Eli Lilly and Company
Indianapolis, IN
USA

Francesca Cunningham
Pharmacy Benefits Management, and Center for Medication Safety
US Department of Veterans Affairs
Hines, IL
USA

Jeffrey R. Curtis
Division of Clinical Immunology & Rheumatology
University of Alabama at Birmingham
Birmingham, AL
USA

Gerald J. Dal Pan
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Rebecca Dickinson
School of Health and Community Studies
Leeds Beckett University
Leeds
UK

Rachael L. DiSantostefano
Janssen Research & Development
Titusville, NJ
USA

Nancy A. Dreyer
IQVIA Real-World & Analytic Solutions
Boston, MA
University of North Carolina at Chapel Hill
Chapel Hill, NC
USA

Scott Evans
Biostatistics Center
The George Washington University
Rockville, MD
USA

Stephen J.W. Evans
Department of Medical Statistics
London School of Hygiene and Tropical Medicine
London
UK

James A. Feinstein
Adult & Child Health Consortium for Health Outcomes Research and Delivery Science
University of Colorado School of Medicine
Aurora, CO
USA

Chris Feudtner
Center for Pediatric Clinical Effectiveness and PolicyLab
The Children's Hospital of Philadelphia
Philadelphia, PA
USA

Brian T. Fisher
Division of Infectious Diseases
The Children's Hospital of Philadelphia
Philadelphia, PA
USA

James H. Flory
Memorial Sloan Kettering Cancer Center
New York, NY
USA

Joshua J. Gagne
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital and Harvard Medical School
Boston, MA, USA

Nicolle M. Gatto
Department of Epidemiology, Worldwide Research and Development
Pfizer, Inc.
New York, NY
USA

Jamie Geier
Department of Epidemiology, Worldwide Research and Development
Pfizer, Inc.
New York, NY
USA

Kate Gelperin
Division of Epidemiology
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Tobias Gerhard
Rutgers Center for Pharmacoepidemiology and Treatment Science
Rutgers Ernest Mario School of Pharmacy
New Brunswick, NJ
USA

Jason M. Glanz
Institute for Health Research
Kaiser Permanente Colorado
and
Department of Epidemiology
Colorado School of Public Health
Denver, CO
USA

Jeremy A. Greene
Department of the History of Medicine
The Johns Hopkins University School of Medicine
Baltimore, MD
USA

Robert Gross
Departments of Medicine (ID) and Epidemiology/Biostatistics/Informatics
Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Thomas P. Gross
Office of Surveillance and Biometrics
Center for Devices and Radiological Health
US Food and Drug Administration
Silver Spring, MD
USA

Sean Hennessy
Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Sonia Hernandez-Diaz
Department of Epidemiology
Harvard T.H. Chan School of Public Health
Boston, MA
USA

Daniel B. Horton
Department of Pediatrics
Rutgers Robert Wood Johnson Medical School
Rutgers Center for Pharmacoepidemiology and Treatment Science
New Brunswick, NJ
USA

Krista F. Huybrechts
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
USA

Judith K. Jones
PharmaLex, Inc.
Fairfax, VA
University of Michigan School of Public Health Summer Program
Ann Arbor, MI
Georgetown University School of Medicine
Washington, DC
USA

Aaron S. Kesselheim
Program On Regulation, Therapeutics, And Law (PORTAL)
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
USA

Stephen E. Kimmel
Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Elyse Kingery
Degge Group, Ltd., a PharmaLex Company
Fairfax, VA
USA

Olaf Klungel
Division of Pharmacoepidemiology & Clinical Pharmacology
Utrecht Institute for Pharmaceutical Sciences (UIPS)
Utrecht University
Utrecht
The Netherlands

Peter Knapp
Department of Health Sciences and the Hull York Medical School
University of York
York
UK

Cynthia Kornegay
Division of Epidemiology
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Tamar Lasky
Office of Biostatistics and Epidemiology
Center for Biologics Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Julie Lauffenburger
Division of Pharmacoepidemiology and Pharmacoeconomics and Center for Healthcare Delivery Sciences
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
USA

David Lee
Pharmaceuticals and Health Technologies Group
Management Sciences for Health
Arlington, VA
USA

Charles E. Leonard
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Samuel M. Lesko
Northeast Regional Cancer Institute
Scranton, PA
USA

Bennett Levitan
Department of Epidemiology
Janssen Research & Development
Titusville, NJ
USA

James D. Lewis
Department of Medicine
Division of Gastroenterology
Department of Biostatistics, Epidemiology, and Informatics
Perelman School of Medicine University of Pennsylvania
Philadelphia, PA
USA

Peter K. Lindenauer
Center for Quality of Care Research
Baystate Medical Center
Springfield, MA
USA

Marie Lindquist
Uppsala Monitoring Centre
WHO Collaborating Centre for International Drug Monitoring
Uppsala
Sweden

Helene Lipton
School of Medicine and Pharmacy
University of California at San Francisco
San Francisco, CA
USA

Christine Y. Lu
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
Boston, MA
USA

Ana Filipa Macedo
Real-World Evidence Solutions Division
QuintilesIMS
Madrid
Spain

Sumit R. Majumdar (deceased)
Formerly of Department of Medicine
Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta
Canada

George Maldonado
Division of Environmental Health Sciences
School of Public Health
University of Minnesota
Minneapolis, MN
USA

Claudia Manzo
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Danica Marinac-Dabic
Office of Clinical Evidence and Analysis
Center for Devices and Radiological Health
US Food and Drug Administration
Silver Spring, MD
USA

Giampiero Mazzaglia
European Medicines Agency
Amsterdam
The Netherlands

Jana McAninch
Division of Epidemiology
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Kenneth L. Melmon (deceased)
Formerly of Stanford University School of Medicine
Stanford, CA
USA

Lubna Merchant
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Allen A. Mitchell
Slone Epidemiology Center at Boston University
Boston, MA
USA

Jingping Mo
Department of Epidemiology, Worldwide Research and Development
Pfizer, Inc.
New York, NY
USA

Esi M. Morgan
Department of Pediatrics
Division of Rheumatology
Cincinnati Children's Hospital Medical Center
University of Cincinnati College of Medicine
Cincinnati, OH
USA

Yola Moride
Faculty of Pharmacy
Université de Montréal
Montreal, Quebec
Canada
Rutgers Center for Pharmacoepidemiology and Treatment Science
New Brunswick, NJ
USA

Sharon-Lise Normand
Department of Health Care Policy
Harvard Medical School
and
Department of Biostatistics
Harvard T.H. Chan School of Public Health
Boston, MA
USA

Alexis Ogdie
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA
USA

Vicki Osborne
Drug Safety Research Unit
Southampton
UK

Robert W. Platt
Departments of Epidemiology, Biostatistics, and Occupational Health, and of Pediatrics
McGill University
Montreal, Quebec
Canada

Anton Pottegård
Clinical Pharmacology and Pharmacy
Department of Public Health
University of Southern Denmark
Odense
Denmark

Nicole Pratt
Quality Use of Medicines and Pharmacy Research Centre
School of Pharmacy and Medical Sciences
University of South Australia
Adelaide, South Australia
Australia

June Raine
Medicines and Healthcare Products Regulatory Agency
London
UK

Robert F. Reynolds
Department of Epidemiology
Global Medical Value, Evidence and Outcomes
GlaxoSmithKline
Upper Providence, PA
USA

Annika Richterich
Faculty of Arts & Social Sciences
Maastricht University
Maastricht
The Netherlands

Mary Elizabeth Ritchey
RTI Health Solutions
RTI International
Research Triangle Park, NC
USA

Sebastian Schneeweiss
Harvard Medical School
Harvard T.H. Chan School of Public Health
Division of Pharmacoepidemiology
Department of Medicine
Brigham and Women's Hospital
Boston, MA
USA

Kevin A. Schulman
Clinical Excellence Research Center
Department of Medicine
Stanford University
Stanford, CA
USA

Alex Secora
Division of Epidemiology
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Art Sedrakyan
Weill Cornell Medical College
New York, NY
USA

Hanna M. Seidling
Division of General Internal Medicine and Primary Care
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
USA
Department of Clinical Pharmacology and Pharmacoepidemiology
Cooperation Unit Clinical Pharmacy
University Hospital Heidelberg
Heidelberg
Germany

Soko Setoguchi
Rutgers Robert Wood Johnson Medical School
Rutgers Center for Pharmacoepidemiology and Treatment Science
New Brunswick, NJ
USA

Saad A.W. Shakir
Drug Safety Research Unit
Southampton
UK

Tom T. Shimabukuro
Immunization Safety Office
Division of Healthcare Quality Promotion
Centers for Disease Control and Prevention
Atlanta, GA
USA

Rachel E. Sobel
Department of Epidemiology, Worldwide Research and Development
Pfizer, Inc.
New York, NY
USA

Stephen B. Soumerai
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
Boston, MA
USA

Tjeerd P. van Staa
Health eResearch Centre
University of Manchester
Manchester
UK

Judy Staffa
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

Brian L. Strom
Rutgers Center for Pharmacoepidemiology and Treatment Science
Rutgers Biomedical and Health Sciences
Newark, NJ
USA

Samy Suissa
McGill University
Centre for Clinical Epidemiology
Lady Davis Research Institute – Jewish General Hospital
Montreal, Quebec
Canada

Janet Sultana
Department of Biomedical and Dental Sciences and Morphofunctional Imaging
University of Messina
Messina
Italy

Sengwee Toh
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
Boston, MA
USA

Gianluca Trifirò
Department of Biomedical and Dental Sciences and Morphofunctional Imaging
University of Messina
Messina
Italy

Shinobu Uzu
Pharmaceuticals and Medical Devices Agency
Tokyo
Japan

Priscilla Velentgas
IQVIA, Real-World Insights
Cambridge, MA
USA

Vera Vlahović-Palčevski
Department of Clinical Pharmacology
University Hospital Rijeka
Department of Pharmacology
University of Rijeka Medical Faculty
Department of Basic Medical Sciences
Faculty of Health Studies
Rijeka
Croatia

Kerstin N. Vokinger
Program On Regulation, Therapeutics, And Law (PORTAL)
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital
Harvard Medical School
Boston, MA
USA

Suzanne L. West
RTI International
Research Triangle Park, NC
Department of Epidemiology
Gillings School of Global Public Health
University of North Carolina
Chapel Hill, NC
USA

Björn Wettermark
Division of Clinical Pharmacology
Department of Laboratory Medicine
Centre for Pharmacoepidemiology
Karolinska Institute
Karolinska University Hospital-Huddinge
Stockholm
Sweden

Ian Chi Kei Wong
Department of Pharmacology and Pharmacy
University of Hong Kong
Hong Kong
UCL School of Pharmacy
London
UK

H. Amy Xia
Amgen, Inc.
Thousand Oaks, CA
USA

Preface

> If the whole materia medica, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind, and all the worse for the fishes.
>
> Oliver Wendell Holmes
> *Comments and Counter-Currents in Medical Science*

The history of drug regulation in the United States is largely a history of political responses to epidemics of adverse drug reactions, each adverse reaction of sufficient public health importance to lead to political pressure for regulatory change.

The initial law, the Pure Food and Drug Act, was passed in 1906. It was a response to the excessive adulteration and misbranding of foods and drugs. The 1938 Food, Drug, and Cosmetic Act was passed in reaction to an epidemic of renal failure resulting from a brand of elixir of sulfanilamide formulated with diethylene glycol. The 1962 Kefauver–Harris Amendment to the Food, Drug, and Cosmetic Act was enacted in response to the infamous "thalidomide disaster," in which children exposed to thalidomide *in utero* were born with phocomelia; that is, with flippers instead of limbs. The resulting regulatory changes led, in part, to the accelerated development of the field of clinical pharmacology, which is the study of the effects of drugs in humans.

Subsequent decades continued to see an accelerating series of accusations about major adverse events possibly associated with drugs. Those discussed in the first edition of this book included liver disease caused by benoxaprofen, subacute myelo-optic-neuropathy (SMON) caused by clioquinol, oculomucocutaneous syndrome caused by practolol, acute flank pain and renal failure caused by suprofen, liver disease caused by ticrynafen, and anaphylactoid reactions caused by zomepirac. Added in the second edition were cardiac arrhythmias from astemizole and terfenadine; hypertension, seizures, and strokes from postpartum use of bromocriptine; deaths from fenoterol; suicidal ideation from fluoxetine; hypoglycemia from human insulin; birth defects from isotretinoin; cancer from depot-medroxyprogesterone; multiple illnesses from silicone breast implants; memory and other central nervous system disturbances from triazolam; and hemolytic anemia and other adverse reactions from temafloxacin. Further added in the third edition were liver toxicity from amoxicillin-clavulanic acid; liver toxicity from bromfenac; cancer and myocardial infarction from calcium channel blockers; cardiac arrhythmias with cisapride; primary pulmonary hypertension and cardiac valvular disease from dexfenfluramine and fenfluramine; gastrointestinal bleeding, postoperative bleeding, deaths, and many other adverse reactions associated with ketorolac; multiple drug interactions with mibefradil; thrombosis from newer oral contraceptives; myocardial infarction from sildenafil; seizures with tramadol; eosinophilia myalgia from tryptophan; anaphylactic

reactions from vitamin K; and liver toxicity from troglitazone. Added in the fourth edition were ischemic colitis from alosetron; myocardial infarction from celecoxib, naproxen, and rofecoxib; rhabdomyolysis from cerivastatin; cardiac arrhythmias from grepafloxacin; stroke from phenylpropanolamine; bronchospasm from rapacuronium; and many others. Added in the fifth edition were progressive multifocal leukoencephalopathy from natalizumab; hepatotoxicity from pamoline and from lumiracoxib; serious cardiovascular complications from rosiglitazone, tegaserod, sibutramine, rimonabant, valdecoxib, pergolide, and propoxyphene; fatal adverse reactions when used with alcohol from palladone; and serious and sometimes fatal brain infections from efalizumab. New in the sixth edition are serious infections of the genital area from sodium-glucose Cotransporter-2 (SGLT2) inhibitors; serious low blood sugar levels and mental health side effects from fluoroquinolones; increased risk of heart-related death and death from all causes from gout medicine febuxostat; increased risk of leg and foot amputations from canagliflozin; possible increased risk of bladder cancer from pioglitazone; heart failure risk from saxagliptin and alogliptin; possible increased risk of heart attack and stroke from testosterone; and potentially fatal heart rhythms from azithromycin. Some of these resulted in drug withdrawals. Published data also suggest that adverse drug reactions could be as much as the fourth leading cause of death. These and other serious but uncommon drug effects have led to the development of new methods to study drug effects in large populations. Academic investigators, the pharmaceutical industry, regulatory agencies, and the legal profession have turned for these methods to the field of epidemiology, the study of the distribution and determinants of disease in populations.

Major new changes have been made in drug regulation and organization, largely in response to a series of accusations about myocardial infarction and stroke caused by analgesics, each detected in long-term prevention trials rather than in normal use of the drugs. For example, the pharmacoepidemiology group at the US Food and Drug Administration (FDA) was doubled in size; the FDA was given new regulatory authority after drug marketing, and was also charged with developing the Sentinel Initiative, a program to conduct medical product safety surveillance in a population to exceed 100 million. Further, the development since January 1, 2006 of Medicare Part D, a US federal program to subsidize prescription drugs for Medicare recipients, introduces to pharmacoepidemiology a new database with a stable population of 25 million, as well as the interest of what may be the largest healthcare system in the world. These developments have brought about major changes for our field.

The bridging of the fields of clinical pharmacology and epidemiology resulted in the development of a new field: pharmacoepidemiology, the study of the use of and effects of drugs in large numbers of people. Pharmacoepidemiology applies the methods of epidemiology to the content area of clinical pharmacology. This new field became the science underlying postmarketing drug surveillance, studies of drug effects that are performed after a drug has been released to the market. In recent years, pharmacoepidemiology has expanded to include many other types of studies as well.

The field of pharmacoepidemiology has grown enormously since the publication of the first edition of this book. The International Society of Pharmacoepidemiology, an early idea when the first edition was written, has grown into a major international scientific force, with over 1476 members from 63 countries, an extremely successful annual meeting attracting more than 1800 attendees, a large number of very active committees and special interest groups, and its own journal. In addition, a number of established journals have targeted pharmacoepidemiology manuscripts as desirable.

As new scientific developments occur within mainstream epidemiology, they are rapidly adopted, applied, and advanced within our field too. We have also become institutionalized as a subfield within the field of clinical pharmacology, with a Drug Utilization and Outcomes community within the American Society for Clinical Pharmacology and Therapeutics, and with pharmacoepidemiology a required part of the clinical pharmacology board examination.

Most of the major international pharmaceutical companies have founded dedicated units to organize and lead their efforts in pharmacoepidemiology, pharmacoeconomics, and quality-of-life studies. The continuing parade of drug safety crises continues to emphasize the need for the field, and some foresighted manufacturers have begun to perform "prophylactic" pharmacoepidemiology studies, so as to have data in hand and available when questions arise, rather than waiting to begin collecting data after a crisis has developed. Pharmacoepidemiologic data are now routinely used for regulatory decisions, and many governmental agencies have been developing and expanding their own pharmacoepidemiology programs. Risk management programs are now required by regulatory bodies with the marketing of new drugs, as a means of improving drugs' benefit/risk balance. Requirements that a drug be proven to be cost-effective have been added to national, local, and insurance healthcare systems, either to justify reimbursement or even to justify drug availability. A number of schools of medicine, pharmacy, and public health have established research programs in pharmacoepidemiology, and a few of them have also established pharmacoepidemiology training programs in response to a desperate need for a bigger pharmacoepidemiology labor force. Pharmacoepidemiologic research funding is now more plentiful, and even support for training is now available, albeit limited.

In the United States, drug utilization review programs are required, by law, of each of the 50 state Medicaid programs, and have been implemented as well in many managed care organizations. However now, years later, the utility of drug utilization review programs has been questioned. In addition, the Joint Commission currently requires that every hospital in the US has an adverse drug reaction monitoring program and a drug use evaluation program, turning every hospital into a mini-pharmacoepidemiology laboratory. Stimulated in part by the interests of the World Health Organization and the Rockefeller Foundation, there is even substantial interest in pharmacoepidemiology in the developing world. Yet, throughout the world, the public's increased concern about privacy has made pharmacoepidemiologic research much more difficult.

In the first edition of this book, the goal was to help introduce this new field to the scientific world. The explosion in interest in the area, the rapid scientific progress that has been made, and the unexpectedly good sales of the first edition led to the second. The continued maturation of what used to be a novel field, the marked increase in sales of the second edition over the first, and the many requests from people all over the world led to the third edition. Thereafter, much in the field has changed, and the fourth edition was prepared. We also produced a textbook version, which has been widely used. Now, seven years after the fifth edition, the field continues to rapidly change, so it is time for a new edition.

In the process, most chapters in the new edition have been thoroughly revised. New chapters have been added, along with many fresh authors. With reorganization of some sections and careful pruning of old chapters, the net size of the book has been kept the same.

As in earlier editions, Part I provides background information on what is included in the field of pharmacoepidemiology, a description of the study designs it uses, a consideration of its unique problem – the requirement for very large sample sizes – and a discussion about when one would want to perform a pharmacoepidemiology

study. Also included is a chapter providing basic principles of clinical pharmacology. Part II presents a series of discussions on the need for the field, the contributions it can make, and some of its problems, from the perspectives of academia, industry, and regulatory agencies. Part III describes the systems that have been developed to perform pharmacoepidemiologic studies, and how each approaches the problem of gathering large sample sizes of study subjects in a cost-effective manner. We no longer attempt to include all the databases in the field, as they have continued to multiply. Instead, in this edition we have combined databases into categories, rather than dedicating a separate chapter to each. Part IV describes selected special opportunities for the application of pharmacoepidemiology to address major issues of importance. These are of particular interest as the field continues to turn its attention to questions beyond just those of adverse drug reactions. Part V presents state-of-the-art discussions of some particular methodologic issues that have arisen in the field. Finally, Part VI provides our personal speculations about the future of pharmacoepidemiology.

This book is not intended as a textbook of adverse drug reactions; that is, a compilation of drug-induced problems organized either by drug or by problem. Nor is it intended primarily as a textbook for use in introductory pharmacoepidemiology courses (for which *Textbook of Pharmacoepidemiology* might be more appropriate). Rather, it is intended to elucidate the methods of investigating adverse drug reactions, as well as other questions of drug effects. It is also not intended as a textbook of clinical pharmacology, organized by disease or by drug, or a textbook of epidemiology, but rather as a text describing the overlap between the two fields.

It is our hope that this book can serve both as a useful introduction to pharmacoepidemiology and as a reference source for the growing number of people interested in this field, in academia, in regulatory agencies, in industry, and in the law. It will also hopefully be useful as a reference text for the numerous courses now underway in this subject. We have been excited by the rapid progress and growth that our field has seen, and delighted that this book has played a small role in assisting this. With this new edition, it will document the major changes that have occurred. In the process, we hope that it can continue to serve to assist in the development of pharmacoepidemiology.

Brian L. Strom, MD, MPH
Stephen E. Kimmel, MD, MSCE
Sean Hennessy, PharmD, PhD

Acknowledgments

There are many individuals and institutions to whom we owe thanks for their contributions to our efforts in preparing this book. Mostly, we thank the contributors who wrote the chapters within it. We greatly enjoyed interacting with them and are grateful for all that we have learned in the process. Over the years, our own pharmacoepidemiology work has been supported mostly by grants, contracts, and cooperative agreements from the US government, especially multiple different institutes of the National Institutes of Health, the Agency for Healthcare Research and Quality, the Food and Drug Administration, and the Department of Veterans Affairs, as well as the Patient Centered Outcomes Research Institute. Other significant funders of our work include private foundations and pharmaceutical companies. Also, we would like to thank the University of Pennsylvania Perelman School of Medicine and Rutgers University. While none of this funding was specifically intended to support the development of this book, without the assistance we would not have been able to develop our careers in pharmacoepidemiology. In addition, we would like to thank Wiley, our publisher, for assistance and insights, both in support of this book and in support of the field's journal, *Pharmacoepidemiology and Drug Safety*.

Chris Rowan's contributions to this book were enormous, encompassing both the role of Managing Editor and reviewing all of the chapters, editing them thoughtfully, and posing additional questions and issues for the authors to address. Finally, Jennifer Hardy provided superb help communicating with the authors and preparing the chapters for submission to Wiley.

BLS would like to thank Steve Kimmel and Sean Hennessy for joining him as co-editors in this and the prior edition. They are two very special and talented men. It has been BLS's pleasure to help to train them, now too many years ago, to help them cultivate their own careers, and to see them blossom into star pharmacoepidemiologists in their own right, now extremely effective and successful. It is a wonderful to be able to share with them this book, which has been an important part of BLS's life and career.

BLS would also like to thank his late parents for the support and education that were critical to his being able to be successful in his career. He would like to thank the late Paul D. Stolley, MD, MPH, and the late Kenneth L. Melmon, MD, for their direction, guidance, and inspiration in the formative years of his career. He would like to thank his trainees, from whom he learns at least as much as he teaches. Last, but certainly not least, BLS would like to thank his family – Lani, Shayna, and Jordi – for accepting the time demands of the book, for tolerating his endless hours working at home on it, and for their ever-present love and support.

SEK expresses his sincere gratitude to BLS for his almost 30 years as a mentor and colleague and for the chance to work on this book, to his parents for providing the foundation for all of his work, and to his family – Alison, David, Benjamin, and Jonathan – for all their support and patience while Dad worked on the book.

SH also thanks BLS, his longtime friend and career mentor, and all of his students, mentees, and collaborators. Finally, he thanks his parents, and his family – Kristin, Landis, and Bridget – for their love and support.

Part I

Introduction

1

What Is Pharmacoepidemiology?

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

> *A desire to take medicine is, perhaps, the great feature which distinguishes man from other animals.*
>
> Sir William Osler, 1891

In recent decades, modern medicine has been blessed with a pharmaceutical armamentarium that is much more powerful than it had before. Although this has given healthcare providers the ability to provide better medical care for their patients, it has resulted too in the ability to do much greater harm. It has also generated an enormous number of product liability suits against pharmaceutical manufacturers, some appropriate and others inappropriate. In fact, the history of drug regulation parallels the history of major adverse drug reaction "disasters." Each change in pharmaceutical law was a political reaction to an epidemic of adverse drug reactions. A 1998 study estimated that 100 000 Americans die each year from adverse drug reactions, and 1.5 million US hospitalizations each year result from adverse drug reactions; yet, 20–70% of adverse drug reactions may be preventable [1]. The harm that drugs can cause has also led to the development of the field of pharmacoepidemiology, which is the focus of this book. More recently, the field has expanded its focus to include in addition many issues other than adverse reactions.

To clarify what is, and what is not, included within the discipline of pharmacoepidemiology, this chapter will begin by defining pharmacoepidemiology, differentiating it from other related fields. The history of drug regulation will then be briefly and selectively reviewed, focusing on the US experience as an example, demonstrating how it has led to the development of this new field. Next, the current regulatory process for the approval of new drugs will be outlined, in order to place the use of pharmacoepidemiology and postmarketing drug surveillance into proper perspective. Finally, the potential scientific and clinical contributions of pharmacoepidemiology will be discussed.

Definition of Pharmacoepidemiology

Pharmacoepidemiology is the study of the use of and the effects of drugs in large numbers of people. The term pharmacoepidemiology obviously contains two components: "pharmaco" and "epidemiology." In order to better appreciate

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

and understand what is and what is not included in this new field, it is useful to compare its scope to that of other related fields. The scope of pharmacoepidemiology will first be compared to that of clinical pharmacology, and then to that of epidemiology.

Pharmacoepidemiology versus Clinical Pharmacology

Pharmacology is the study of the effects of drugs. *Clinical pharmacology* is the study of the effects of drugs in humans (see also Chapter 2). Pharmacoepidemiology obviously can be considered, therefore, to fall within clinical pharmacology. In attempting to optimize the use of drugs, one central principle of clinical pharmacology is that therapy should be individualized, or tailored, to the needs of the particular patient at hand. This individualization of therapy requires the determination of a risk/benefit ratio specific to the patient. Doing so requires a prescriber to be aware of the potential beneficial and harmful effects of the drug in question and to know how elements of the patient's clinical status might modify the probability of a good therapeutic outcome. For example, consider a patient with a serious infection, serious liver impairment, and mild impairment of his or her renal function. In considering whether to use gentamicin to treat the infection, it is not sufficient to know that gentamicin has a small probability of causing renal disease. A good clinician should realize that a patient who has impaired liver function is at a greater risk of suffering from this adverse effect than one with normal liver function [2]. Pharmacoepidemiology can be useful in providing information about the beneficial and harmful effects of any drug, thus permitting a better assessment of the risk/benefit balance for the use of any particular drug in any particular patient.

Clinical pharmacology is traditionally divided into two basic areas: pharmacokinetics and pharmacodynamics. *Pharmacokinetics* is the study of the relationship between the dose administered of a drug and the serum or blood level achieved. It deals with drug absorption, distribution, metabolism, and excretion. *Pharmacodynamics* is the study of the relationship between drug level and drug effect. Together, these two fields allow one to predict the effect one might observe in a patient from administering a certain drug regimen. Pharmacoepidemiology encompasses elements of both of these fields, exploring the effects achieved by administering a drug regimen. It does not normally involve or require the measurement of drug levels. However, pharmacoepidemiology can be used to shed light on the pharmacokinetics of a drug when used in clinical practice, such as exploring whether aminophylline is more likely to cause nausea when administered to a patient who is simultaneously taking cimetidine. However, to date this is a relatively novel application of the field.

Specifically, the field of pharmacoepidemiology has primarily concerned itself with the study of adverse drug effects. Adverse reactions have traditionally been separated into those which are the result of an exaggerated but otherwise usual pharmacologic effect of the drug, sometimes called *type A reactions*, versus those which are aberrant effects, so called *type B reactions* [3]. Type A reactions tend to be common, dose-related, predictable, and less serious. They can usually be treated by simply reducing the dose of the drug. They tend to occur in individuals who have one of three characteristics. First, the individuals may have received more of a drug than is customarily required. Second, they may have received a conventional amount of the drug, but they may metabolize or excrete it unusually slowly, leading to drug levels that are too high (see also Chapter 2). Third, they may have normal drug levels, but for some reason are overly sensitive to the drug.

In contrast, type B reactions tend to be uncommon, not related to dose, unpredictable,

and potentially more serious. They usually require cessation of the drug. They may be due to what are known as hypersensitivity reactions or immunologic reactions. Alternatively, type B reactions may be some other idiosyncratic reaction to the drug, either due to some inherited susceptibility (e.g., glucose-6-phosphate dehydrogenase deficiency) or due to some other mechanism. Regardless, type B reactions are the most difficult to predict or even detect, and represent the major focus of many pharmacoepidemiologic studies of adverse drug reactions.

One typical approach to studying adverse drug reactions has been the collection of spontaneous reports of drug-related morbidity or mortality (see Chapter 10), sometimes called pharmacovigilance (although other times that term is used to refer to all of pharmacoepidemiology). However, determining causation in case reports of adverse reactions can be problematic (see Chapter 29), as can attempts to compare the effects of drugs in the same class (see Chapter 26). Further, drug–drug interactions, predicted based on pharmacokinetic data (see Chapter 2), require massive sample sizes to confirm in people (see Chapter 40). This has led academic investigators, industry, the US Food and Drug Administration (FDA), and the legal community to turn to the field of epidemiology. Specifically, *studies of adverse effects* have been supplemented with *studies of adverse events*. In the former, investigators examine case reports of purported adverse drug reactions and attempt to make a subjective clinical judgment on an *individual* basis about whether the adverse outcome was actually caused by the antecedent drug exposure. In the latter, controlled studies are performed examining whether the adverse outcome under study occurs more often in an exposed *population* than in an unexposed population. This marriage of the fields of clinical pharmacology and epidemiology has resulted in the development of a further field: pharmacoepidemiology.

Pharmacoepidemiology versus Epidemiology

Epidemiology is the study of the distribution and determinants of diseases in populations (see Chapter 3). Since pharmacoepidemiology is the study of the use of and effects of drugs in large numbers of people, it obviously falls within epidemiology as well. Epidemiology is also traditionally subdivided into two basic areas. The field began as the study of infectious diseases in large populations; that is, epidemics. It has since been expanded to encompass the study of chronic diseases. The field of pharmacoepidemiology uses the techniques of chronic disease epidemiology to study the use of and the effects of drugs. Although application of the methods of pharmacoepidemiology can be useful in undertaking the clinical trials of drugs that are performed before marketing [4], the major application of these principles is after drug marketing. This has primarily been in the context of postmarketing drug surveillance, although in recent years the interests of pharmacoepidemiologists have broadened considerably. Now, as will be made clearer in future chapters, pharmacoepidemiology is considered of importance in the whole life cycle of a drug, from the time it is first discovered or synthesized through to when it is no longer sold as a drug.

Thus, pharmacoepidemiology is a relatively new applied field, bridging between clinical pharmacology and epidemiology. From clinical pharmacology, pharmacoepidemiology borrows its focus of inquiry. From epidemiology, pharmacoepidemiology borrows its methods of inquiry. In other words, it applies the methods of epidemiology to the content area of clinical pharmacology. In the process, multiple special logistical approaches have been developed and multiple special methodologic issues have arisen. These are the primary foci of this book.

Historical Background

Early Legislation

The history of drug regulation in the US is similar to that in most developed countries, and reflects the growing involvement of governments in attempting to insure that only safe and effective drug products were available and that appropriate manufacturing and marketing practices were used. The initial US law, the Pure Food and Drug Act, was passed in 1906, in response to excessive adulteration and misbranding of the food and drugs available at that time. There were no restrictions on sales or requirements for proof of the efficacy or safety of marketed drugs. Rather, the law simply gave the federal government the power to remove from the market any product that was adulterated or misbranded. The burden of proof was on the federal government.

In 1937, over 100 people died from renal failure as a result of the marketing by the Massengill Company of elixir of sulfanilimide dissolved in diethylene glycol [5]. In response, Congress passed the 1938 Food, Drug, and Cosmetic Act. Preclinical toxicity testing was required for the first time. In addition, manufacturers were required to gather clinical data about drug safety and to submit these data to FDA before drug marketing. The FDA had 60 days to object to marketing or else it would proceed. No proof of efficacy was required.

Little attention was paid to adverse drug reactions until the early 1950s, when it was discovered that chloramphenicol could cause aplastic anemia [6]. In 1952, the first textbook of adverse drug reactions was published [7]. In the same year, the American Medical Association (AMA) Council on Pharmacy and Chemistry established the first official registry of adverse drug effects, to collect cases of drug-induced blood dyscrasias [8]. In 1960, the FDA began to collect reports of adverse drug reactions and sponsored new hospital-based drug-monitoring programs. The Johns Hopkins Hospital and the Boston Collaborative Drug Surveillance Program developed the use of in-hospital monitors to perform cohort studies to explore the short-term effects of drugs used in hospitals [9,10]. This approach was later to be transported to the University of Florida–Shands Teaching Hospital as well [11].

In the winter of 1961, the world experienced the infamous "thalidomide disaster." Thalidomide was marketed as a mild hypnotic, and had no obvious advantage over other drugs in its class. Shortly after its marketing, a dramatic increase was seen in the frequency of a previously rare birth defect, phocomelia: the absence of limbs or parts of limbs, sometimes with the presence instead of flippers [12]. Epidemiologic studies established its cause to be *in utero* exposure to thalidomide. In the UK, this resulted in the establishment in 1968 of the Committee on Safety of Medicines. Later, the World Health Organization (WHO) established a bureau to collect and collate information from this and other similar national drug-monitoring organizations (see Chapter 10).

The US had never permitted the marketing of thalidomide and so was fortunately spared this epidemic. However, the "thalidomide disaster" was so dramatic that it resulted in regulatory change in the US as well. Specifically, in 1962 the Kefauver–Harris Amendments were passed. These amendments strengthened the requirements for proof of drug safety, requiring extensive preclinical pharmacologic and toxicologic testing before a drug could be tested in humans. The data from these studies were required to be submitted to the FDA in an Investigational New Drug (IND) application before clinical studies could begin. Three explicit phases of clinical testing were defined, which are described in more detail later in this chapter. In addition, a new requirement was added to the clinical testing, for "substantial evidence that the drug will have the effect it purports or is represented to have." "Substantial evidence" was defined as "adequate and well-controlled investigations,

including clinical investigations." Functionally, this has generally been interpreted as requiring randomized clinical trials to document drug efficacy before marketing. This new procedure also delayed drug marketing until the FDA explicitly gave approval. With some modifications, these are the requirements still in place in the US today. In addition, the amendments required the review of all drugs approved between 1938 and 1962, to determine if they too were efficacious. The resulting DESI (Drug Efficacy Study Implementation) process, conducted by the National Academy of Sciences' National Research Council with support from a contract from the FDA, was not completed until years later, and resulted in the removal from the US market of many ineffective drugs and drug combinations. The result of all these changes was a great prolongation of the approval process, with attendant increases in the cost of drug development, in what was termed the "drug lag" [13] (discussed later in the chapter). However, the drugs that do reach the market are presumably much safer and more effective.

Drug Crises and Resulting Regulatory Actions

Despite the more stringent process for drug regulation, subsequent years have seen a series of major adverse drug reactions. Subacute myelo-optic-neuropathy (SMON) was found in Japan to be caused by clioquinol, a drug marketed in the early 1930s but not discovered to cause this severe neurologic reaction until 1970 [14]. In the 1970s, clear cell adenocarcinoma of the cervix and vagina and other genital malformations were found to be due to *in utero* exposure to diethylstilbestrol two decades earlier [15]. The mid-1970s saw the UK discovery of the oculomucocutaneous syndrome caused by practolol, five years after drug marketing [16]. In 1980, the drug ticrynafen was noted to cause deaths from liver disease [17]. In 1982, benoxaprofen was noted to do the same [18]. Subsequently the use of zomepirac, another nonsteroidal anti-inflammatory drug, was noted to be associated with an increased risk of anaphylactoid reactions [19]. Serious blood dyscrasias were linked to phenylbutazone [20]. Small intestinal perforations were noted to be caused by a particular slow-release formulation of indomethacin [21]. Bendectin®, a combination product indicated to treat nausea and vomiting in pregnancy, was removed from the market because of litigation claiming it was a teratogen, despite the absence of valid scientific evidence to justify this claim [22] (see Chapter 22). Acute flank pain and reversible acute renal failure were noted to be caused by suprofen [23]. Isotretinoin was almost removed from the US market because of the birth defects it causes [24,25]. The Eosinophilia-Myalgia syndrome was linked to a particular brand of L-tryptophan [26]. Triazolam, thought by The Netherlands in 1979 to be subject to a disproportionate number of central nervous system side effects [27], was discovered by the rest of the world to be problematic in the early 1990s [28–30]. Silicone breast implants, inserted by the millions in the US for cosmetic purposes, were accused of causing cancer, rheumatologic disease, and many other problems, and restricted from use except for breast reconstruction after mastectomy [31]. Human insulin was marketed as one of the first of the new biotechnology drugs, but soon thereafter was accused of causing a disproportionate amount of hypoglycemia [32–36]. Fluoxetine was marketed as a major new, important and commercially successful psychiatric product, but then lost a large part of its market due to accusations about its association with suicidal ideation [37,38]. An epidemic of deaths from asthma in New Zealand was traced to fenoterol [39–41], and later data suggested that similar, although smaller, risks might be present with other beta-agonist inhalers [42]. The possibility was raised of cancer from depot-medroxyprogesterone, resulting in initial refusal to allow its marketing for this purpose in the US

[43], multiple studies [44,45], and ultimate approval. Arrhythmias were linked to the use of the antihistamines terfenadine and astemizole [46,47]. Hypertension, seizures, and strokes were noted from postpartum use of bromocriptine [48,49]. Multiple different adverse reactions were linked to temafloxacin [50]. Other examples include liver toxicity from amoxicillin-clavulanic acid [51]; liver toxicity from bromfenac [52,53]; cancer, myocardial infarction, and gastrointestinal bleeding from calcium channel blockers [54–61]; arrhythmias with cisapride interactions [62–65]; primary pulmonary hypertension and cardiac valvular disease from dexfenfluramine and fenfluramine [66–68]; gastrointestinal bleeding, postoperative bleeding, deaths, and many other adverse reactions associated with ketorolac [69–72]; multiple drug interactions with mibefradil [73]; thrombosis from newer oral contraceptives [74–77]; myocardial infarction from sildenafil [78]; seizures from tramadol [79,80]; anaphylactic reactions from vitamin K [81]; liver toxicity from troglitazone [82–85]; and intussusception from rotavirus vaccine [86].

Later drug crises have occurred due to allegations of ischemic colitis from alosetron [87]; rhabdomyolysis from cerivastatin [88]; bronchospasm from rapacuronium [89]; torsades de pointes from ziprasidone [90]; hemorrhagic stroke from phenylpropanolamine [91]; arthralgia, myalgia, and neurologic conditions from Lyme vaccine [92]; multiple joint and other symptoms from anthrax vaccine [93]; myocarditis and myocardial infarction from smallpox vaccine [94]; and heart attack and stroke from rofecoxib [95].

Major adverse drug reactions continue to plague new drugs, and in fact have been as common if not more common in the last several decades. In total, 36 different oral prescription drug products have been removed from the US market since 1980 alone (alosetron, 2000; aprotinin, 2007; astemizole, 1999; benoxaprofen, 1982; bromfenac, 1998; cerivastatin, 2001; cisapride, 2000; dexfenfluramine, 1997; efalizumab, 2009; encainide, 1991; etretinate, 1998; fenfluramine, 1998; flosequinan, 1993; grepafloxin, 1999; levomethadyl, 2003; lumiracoxib, 2007; mibefradil, 1998; natalizumab, 2005; nomifensine, 1986; palladone, 2005; pemoline, 2005; pergolide, 2010; phenylpropanolamine, 2000; propoxyphene, 2010; rapacuronium, 2001; rimonabant, 2010; rofecoxib, 2004; sibutramine, 2010; suprofen, 1987; tegaserod, 2007; terfenadine, 1998; temafloxacin, 1992; ticrynafen, 1980; troglitazone, 2000; valdecoxib, 2007; zomepirac, 1983). The licensed vaccines against rotavirus [86] and Lyme [92] were also withdrawn because of safety concerns (see Chapter 20). Further, between 1990 and 2004, at least 15 noncardiac drugs, including astemizole, cisapride, droperidol, grepafloxacin, halofantrine, pimozide, propoxyphene, rofecoxib, sertindole, sibutramine, terfenadine, terodiline, thioridazine, vevacetylmethadol, and ziprasidone, were subject to significant regulatory actions because of cardiac concerns [96].

Since 1993, trying to deal with drug safety problems, the FDA morphed its extant spontaneous reporting system into the MedWatch program of collecting spontaneous reports of adverse reactions (see Chapters 8 and 10), as part of that issuing monthly notifications of label changes. Compared to the 20–25 safety-related label changes that were being made every month by mid-1999, between 19 and 57 safety-related label changes (boxed warnings, warnings, contraindications, precautions, adverse events) were made every month in 2009 [97]. From January of 2010 to July of 2016, there were 3324 safety-related label changes, with a range per month of 19–87 (median 41). Among all safety-related label changes (January 2010 to July 2016), 8%, 13%, 56%, and 65% were boxed warnings, contraindications, warnings, and precautions, respectively [97].

According to a study from a number of years ago by the US Government Accountability Office, 51% of approved drugs have serious

adverse effects not detected before approval [98]. Further, there is recognition that the initial dose recommended for a newly marketed drug is often incorrect, and needs monitoring and modification after marketing [99–101].

In some of the examples given, the drug was never convincingly linked to the adverse reaction, yet many of these accusations led to the removal of the drug involved from the market. Interestingly, however, this withdrawal was not necessarily performed in all of the different countries in which each drug was marketed. Most of these discoveries have led to litigation as well, and a few have even resulted in criminal charges against the pharmaceutical manufacturer and/or some of its employees (see Chapter 9).

Legislative Actions Resulting from Drug Crises

Through the 1980s, there was concern that an underfunded FDA was approving drugs too slowly, and that the US suffered, compared to Europe, from a "drug lag" [102]. To provide additional resources to the FDA to help expedite the drug review and approval process, Congress passed in 1992 the Prescription Drug User Fee Act (PDUFA), allowing the FDA to charge manufacturers a fee for reviewing new drug applications [103,104]. This legislation was reauthorized by Congress three more times: PDUFA II, also called the Food and Drug Modernization Act of 1997; PDUFA III, also called the Public Health Security and Bioterrorism Preparedness and Response Act of 2002; and PDUFA IV, also called the Food and Drug Administration Amendments (FDAAA-PL 110-85) of 2007. The goals for PDUFA I–IV were to enable the FDA to complete review of over 90% of priority drug applications in 6 months, and complete review of over 90% of standard drug applications in 12 months (under PDUFA I) or 10 months (under PDUFA II–IV). In addition to reauthorizing the collection of user fees from the pharmaceutical industry, PDUFA II allowed the FDA to accept a single well-controlled clinical study under certain conditions, to reduce drug development time. The result was a system where more than 550 new drugs were approved by the FDA in the 1990s [105].

However, whereas 1400 FDA employees in 1998 worked with the drug approval process, only 52 monitored safety; FDA spent a mere $2.4 million on extramural safety research. This state of affairs coincided with the growing numbers of drug crises already cited. With successive reauthorizations of PDUFA, this markedly changed. PDUFA III for the first time allowed the FDA to use a small portion of the user fees for postmarketing drug safety monitoring, to address safety concerns.

Nevertheless, there now was growing concern, in Congress and among the US public, that perhaps the FDA was approving drugs too *fast* [106,107]. There were also calls for the development of an independent drug safety board, analogous to the National Transportation Safety Board [108,109], with a mission much wider than the FDA's regulatory mission, to complement the latter. For example, such a board could investigate drug safety crises such as those discussed, looking for ways to prevent them, and could deal with issues such as improper physician use of drugs, the need for training, and the development of new approaches to the field of pharmacoepidemiology.

Recurrent concerns about the FDA's management of postmarketing drug safety issues led to a systematic review of the entire drug risk assessment process. In 2006, the US General Accountability Office issued its report of a review of the organizational structure and effectiveness of FDA's postmarketing drug safety decision-making [100], followed in 2007 by the Institute of Medicine's independent assessment [110]. Important weaknesses were noted in the current system, including failure of the FDA's Office of New Drugs and Office of Drug Safety to communicate with each other on safety issues, failure of the FDA to track ongoing postmarketing studies, the ambiguous role of the

FDA's Office of Drug Safety in scientific advisory committees, limited authority by the FDA to require the pharmaceutical industry to perform studies to obtain needed data, concerns about culture problems at the FDA where recommendations by members of its drug safety staff were not followed, and concerns about conflicts of interest involving advisory committee members. This Institute of Medicine report was influential in shaping PDUFA IV.

Indeed, with the passage of those amendments, the FDA's authority was substantially increased, with the ability, for example, to require postmarketing studies and levy heavy fines if these requirements were not met. Further, its resources were substantially increased, with a specific charge to (i) fund epidemiology best practices and data acquisition ($7 million in fiscal 2008, increasing to $9.5 million in fiscal 2012); (ii) fund new drug trade name review ($5.3 million in fiscal 2008, rising to $6.5 million in fiscal 2012); and (iii) fund risk management and communication ($4 million in fiscal 2008, rising to $5 million in fiscal 2012) [111] (see also Chapter 24). In addition, in another use of the new PDUFA funds, the FDA plans to develop and implement agency-wide and special-purpose postmarket information technology (IT) systems, including the MedWatch Plus Portal, the FDA Adverse Event Reporting System, the Sentinel System (a virtual national medical product safety system; see Chapter 25), and the Phonetic and Orthographic Computer Analysis System to find similarities in spelling or sound between proposed proprietary drug names that might increase the risk of confusion and medication errors [111].

The Food and Drug Administration Safety and Innovation Act of 2012 (FDASIA), the fifth authorization of PDUFA, expanded the FDA's authority with the ability to safeguard and advance public health by: (i) "giving the authority to collect user fees from industry to fund reviews of innovator drugs, medical devices, generic drugs and biosimilar biological products"; (ii) "promoting innovation to speed patient access to safe and effective products"; (iii) "increasing stakeholder involvement in FDA processes"; and (iv) "enhancing the safety of the drug supply chain" [112]. Also enacted in 2012, the Generic Drug User Fee Amendments (GDUFA) permitted the FDA to assess industry user fees with the intention of increasing the predictability and timeliness of generic drug application reviews [113]. The Biosimilar User Fee Act (BsUFA), also enacted in 2012, authorized the FDA to collect fees directly from biosimilar drug product applicants to expedite the review of biosimilar applications [114]. The FDA Reauthorization Act of 2017 (FDARA) reauthorized PDUFA, GDUFA, and BsUFA through fiscal year 2022.

Among other aims, the 21st Century Cures Act (enacted in December 2016) was intended to expedite the process by which new drugs and devices are approved by easing the requirements put on drug companies looking for FDA approval on new products or new indications on existing drugs. It calls for the use of "data summaries" to support the approval of certain drugs for new indications, rather than full clinical trial data. It also allows drug companies to promote off-label uses to insurance companies, enabling them to expand their markets. Of particular relevance to pharmacoepidemiology, it permits the use of "real world evidence" rather than just clinical trial results [115]. Depending on how these new rules are interpreted, this could massively change drug development in the US, and in particular the role of pharmacoepidemiology in that drug development.

Intellectual Development of Pharmacoepidemiology Emerging from Drug Crises

Several developments in the 1960s can be thought to have marked the beginning of the field of pharmacoepidemiology. The Kefauver–Harris Amendments that were introduced in

1962 required formal safety studies for new drug applications. The DESI program that was undertaken by the FDA as part of those amendments required formal efficacy studies for old drugs that were approved earlier. These requirements created a demand for new expertise and new methods. In addition, the mid-1960s saw the publication of a series of drug utilization studies [116–120]. These provided the first descriptive information on how physicians use drugs, and began a series of investigations of the frequency and determinants of poor prescribing (see also Chapters 18 and 19).

In part in response to concerns about adverse drug effects, the early 1970s saw the development of the Drug Epidemiology Unit, now the Slone Epidemiology Center, which extended the hospital-based approach of the Boston Collaborative Drug Surveillance Program by collecting lifetime drug exposure histories from hospitalized patients and using these to perform hospital-based case–control studies [121] (see Chapter 16). The year 1976 saw the formation of the Joint Commission on Prescription Drug Use, an interdisciplinary committee of experts charged with reviewing the state of the art of pharmacoepidemiology at that time, as well as providing recommendations for the future [122]. The Computerized Online Medicaid Analysis and Surveillance System (COMPASS®) was first developed in 1977, using Medicaid billing data to perform pharmacoepidemiologic studies [123] (see Chapter 12). The Drug Surveillance Research Unit, now called the Drug Safety Research Trust, was developed in the UK in 1980, with its innovative system of prescription event monitoring [124] (see Chapter 15). Each of these represented major contributions to the field of pharmacoepidemiology, and together with newer approaches are reviewed in Part III of this book.

In the examples of drug crises mentioned earlier, there were serious but uncommon drug effects, and these experiences led to an accelerated search for new methods to study drug effects in large numbers of patients. This resulted in a shift from adverse effect studies to adverse event studies, with a concomitant increasing use of new data resources and new methods to study adverse reactions. The American Society for Clinical Pharmacology and Therapeutics issued, in 1990, a position paper on the use of purported postmarketing drug surveillance studies for promotional purposes [125], and the International Society for Pharmacoepidemiology (ISPE) issued, in 1996, Guidelines for Good Epidemiology Practices for Drug, Device, and Vaccine Research in the United States [126], which were updated in 2007 [127] and 2015. Since the late 1990s, pharmacoepidemiologic research has also been increasingly burdened by concerns about patient confidentiality [128–132] (see also Chapter 31).

There is also increasing recognition that most of the risk from most drugs to most patients occurs from known reactions to old drugs. As an attempt to address concerns about underuse, overuse, and adverse events of medical products and medical errors that may cause serious impairment to patient health, a new program of Centers for Education and Research on Therapeutics (CERTs) was authorized under the FDA Modernization Act of 1997 (as part of the same legislation that reauthorized PDUFA II). Starting in 1999 and incrementally adding more centers in 2002, 2006, and 2007, the Agency for Healthcare Research and Quality (AHRQ) that was selected to administer this program had funded up to 14 CERTs [133], although the program ended in 2016 (see also Chapter 6).

The research and education activities sponsored by AHRQ through the CERTs program since the late 1990s take place in academic centers. The CERTs conduct research on therapeutics, exploring new uses of drugs, ways to improve the effective uses of drugs, and the risks associated with new uses or combinations of drugs. They also develop educational modules and materials for disseminating the research

findings about medical products. With the development of direct-to-consumer advertising of drugs since the mid-1980s in the US, the CERTs' role in educating the public and healthcare professionals by providing evidence-based information has become especially important.

Another impetus for research on drugs resulted from one of the mandates (in Sec. 1013) of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 to provide beneficiaries with scientific information on the outcomes, comparative clinical effectiveness, and appropriateness of healthcare items and services [134]. In response, the AHRQ created in 2005 the DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Network to support in academic settings the conduct of studies on the effectiveness, safety, and usefulness of drugs and other treatments and services [135]. This too ended, in 2012.

Another major new initiative of close relevance to pharmacoepidemiology is risk management. There is increasing recognition that the risk/benefit balance of some drugs can only be considered acceptable with active management of their use, to maximize their efficacy and/or minimize their risk. In response, in the late 1990s, new initiatives ranged from FDA requirements for risk management plans to an FDA Drug Safety and Risk Management Advisory Committee, and the issuing of risk minimization and management guidance in 2005. More information is provided in Chapters 8 and 24.

Another initiative closely related to pharmacoepidemiology is the Patient Safety movement. In the Institute of Medicine's report, "To Err Is Human: Building a Safer Health System," the authors note that (i) "even apparently single events or errors are due most often to the convergence of multiple contributing factors"; (ii) "preventing errors and improving safety for patients requires a systems approach in order to modify the conditions that contribute to errors"; and (iii) "the problem is not bad people; the problem is that the system needs to be made safer" [136]. In this framework, the concern is not about substandard or negligent care, but rather about errors made by even the best trained, brightest, and most competent professional health caregivers and/or patients. From this perspective, the important research questions ask about the conditions under which people make errors, the types of errors being made, and the types of systems that can be put into place to prevent errors altogether when possible. Errors that are not prevented must be identified and corrected efficiently and quickly, before they inflict harm. Turning specifically to medications, from 2.4% to 6.5% of hospitalized patients suffer adverse drug effects, prolonging hospital stays by 2 days, and increasing costs by $2000–2600 per patient [137–140]. Over 7000 US deaths were attributed to medication errors in 1993 [141]. Although these estimates have been disputed [142–147], the overall importance of reducing these errors has not been questioned. In recognition of this problem, the AHRQ launched a major new grant program of over 100 projects at its peak with over $50 million/year of funding. While only a portion of this is dedicated to medication errors, they are clearly a focus of interest and relevance to many. More information is provided in Chapter 41.

The 1990s and especially the 2000s saw another shift in the field, away from its exclusive emphasis on drug utilization and adverse reactions, to the inclusion of other interests as well, such as the use of pharmacoepidemiology to study beneficial drug effects, the application of health economics to the study of drug effects, quality-of-life studies, meta-analysis, studies of biologics, data mining, studies of drugs of abuse, drug interactions, and so on. These new foci are discussed in more detail in Parts IV and V of this book.

Moreover, with the publication of the results from the Women's Health Initiative indicating that combination hormone replacement therapy

causes an increased risk of myocardial infarction rather than a decreased risk [148,149], there has been increased concern about reliance solely on nonexperimental methods to study drug safety after marketing [150–153]. This led to increased use of massive randomized clinical trials as part of postmarketing surveillance (see Chapter 32). This is especially important because often the surrogate markers used for drug development cannot necessarily be relied upon to map completely to true clinical outcomes [154].

Finally, with the advent of the Obama administration in the US, there was enormous interest in comparative effectiveness research (CER). CER was defined in 2009 by the Federal Coordinating Council for Comparative Effectiveness Research as "the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in "real world" settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances" [155]. By this definition, CER includes three key elements: (i) evidence synthesis, evidence generation, and evidence dissemination. Typically, CER is conducted through observational studies of either large administrative or medical record databases (see Part IIIb), or large naturalistic clinical trials (see Chapter 32). In many ways, the UK has been focusing on CER for years via its National Institute for Health and Clinical Excellence (NICE), an independent organization responsible for providing national guidance on promoting good health and preventing and treating ill health [156]. However, the Obama administration included $1.1 billion for CER in its federal stimulus package, and had plans for hundreds of millions of dollars of support per year thereafter. While CER does not overlap completely with pharmacoepidemiology, the scientific approaches are very close. Pharmacoepidemiologists evaluate the use and effects of medications. CER investigators compare, in the real world, the safety and benefits of one treatment to those of another. CER extends beyond pharmacoepidemiology in that it can include more than just drugs; pharmacoepidemiology extends beyond CER in that it includes studies comparing exposed to unexposed patients, not just alternative exposures. However, to date, most work done in CER has been in pharmacoepidemiology. See Chapter 26 for more discussion.

The Current Drug Approval Process

Drug Approval in the US

Until the early 1990s, there was a decline in the number of novel drugs approved per year [101,157], while the cost of bringing a drug to market has risen sharply [158]. The total cost of drug development to the pharmaceutical industry increased from $24 billion in 1999, to $32 billion in 2002 [159], and to $65.2 billion on research and development in 2008 [160]. The cost to discover and develop a drug that successfully reached the market rose from over $800 million in 2004 [161] to an estimated $1.3–1.7 billion currently [162]. In addition to the sizable costs of research and development, a substantial part of the total cost is determined also by the regulatory requirement to test new drugs during several premarketing and postmarketing phases, as will be reviewed next.

The current drug approval process in the US and most other developed countries includes preclinical animal testing followed by three phases of clinical testing. Phase I testing is usually conducted in just a few normal volunteers, and represents the initial trials of the drug in

humans. Phase I trials are generally conducted by clinical pharmacologists, to determine the metabolism of the drug in humans, a safe dosage range in humans, and to exclude any extremely common toxic reactions that are unique to humans.

Phase II testing is also generally conducted by clinical pharmacologists, on a small number of patients who have the target disease. Phase II testing is usually the first time patients are exposed to the drug. Exceptions are drugs that are so toxic that it would not normally be considered ethical to expose healthy individuals to them, like cytotoxic drugs. For these, patients are used for Phase I testing as well. The goals of Phase II testing are to obtain more information on the pharmacokinetics of the drug and on any relatively common adverse reactions, and to obtain initial information on its possible efficacy. Specifically, Phase II is used to determine the daily dosage and regimen to be tested more rigorously in Phase III.

Phase III testing is performed by clinician-investigators in a much larger number of patients, in order to rigorously evaluate the drug's efficacy and provide more information on its toxicity. At least one of the Phase III studies needs to be a randomized clinical trial (see Chapter 3). To meet FDA standards, at least one of the randomized clinical trials usually needs to be conducted in the US. Generally between 500 and 3000 patients are exposed to a drug during Phase III, even if drug efficacy can be demonstrated with much smaller numbers, in order to be able to detect less common adverse reactions. For example, a study including 3000 patients would allow one to be 95% certain of detecting any adverse reactions that occur in at least 1 exposed patient out of 1000. At the other extreme, a total of 500 patients would allow one to be 95% certain of detecting any adverse reactions that occur in 6 or more patients out of every 1000 exposed. Adverse reactions that occur less commonly than these are less likely to be detected in these premarketing studies. The sample sizes needed to detect drug effects are discussed in more detail in Chapter 4. Nowadays, with the increased focus on drug safety, premarketing dossiers are sometimes being extended well beyond 3000 patients. However, as one can tell from the sample size calculations in Chapter 4 and Appendix A, by itself these larger numbers lead to little additional information being gained about adverse drug reactions, unless one were to increase to perhaps 30 000 patients, well beyond the scope of most premarketing studies.

Finally, Phase IV testing is the evaluation of the effects of drugs after general marketing. The bulk of this book is devoted to such efforts.

Drug Approval in Other Countries

Outside the US, national systems for the regulation and approval of new drugs vary greatly, even among developed countries and especially between developed and developing countries. While in most developed countries at least the general process of drug development is very analogous to that in the US, its implementation varies widely. A WHO comparative analysis of drug regulation in 10 countries found that not all even have a written national drug policy document [163]. Regulation of medicines in some countries is centralized in a single agency that performs the gamut of functions, involving product registration, licensing, product review, approval for clinical trials, postmarketing surveillance, and inspection of manufacturing practice. Examples for this are Health Canada [164], the China Food and Drug Administration (CFDA) [165], the Medicines Agency in Denmark [166], the Medicines Agency in Norway [167], the Center for Drug Administration in Singapore [168], and the Medicines and Medical Devices Safety Authority in New Zealand [169]. In other countries, regulatory functions are distributed among different agencies. An example of the latter is The Netherlands, where the Ministry of Health,

Welfare and Sports performs the functions of licensing; the Healthcare Inspectorate checks on general manufacturing practice; and the Medicines Evaluation Board performs the functions of product assessment and registration and adverse drug reaction monitoring [163]. As another example, in Singapore two independent agencies (the Center for Pharmaceutical Administration and the Center for Drug Evaluation) were previously responsible for medicinal regulation and evaluation, but are currently merged into a single agency (the Center for Drug Administration) [168].

Another dimension on which countries may vary is the degree of autonomy of regulatory decisions from political influence. Drug regulation in most countries is performed by a department within the executive branch (Australia, Cuba, Cyprus, Tunisia, and Venezuela are examples cited by the WHO report, and Denmark [166], India [170], and New Zealand [169] are other examples). In other countries, this function is performed by an independent commission or board. An example of the latter arrangement is The Netherlands, where members of the Medicines Evaluation Board are appointed directly by the Crown, thereby enabling actions that are independent of interference by other government authorities, such as the Ministry of Health [163]. All 10 countries examined by the WHO require registration of pharmaceutical products, but they differ on the documentation requirements for evidence of safety and efficacy [163]. Some countries carry out independent assessments while others, especially many developing countries, rely on WHO assessments or other sources [163]. With the exception of Cyprus, the remaining nine countries surveyed by the WHO were found to regulate the conduct of clinical trials, but with varying rates of participation of healthcare professionals in reporting adverse drug reactions [163]. Another source noted that countries also differ on the extent of emphasis on quantitative or qualitative analysis for assessing pre- and postmarketing data [171].

Further, within Europe, each country has its own regulatory agency, for instance the UK Medicines and Healthcare Products Regulatory Agency (MHRA), formed in 2003 as a merger of the Medicines Control Agency (MCA) and the Medical Devices Agency (MDA). In addition, since January 1998, some drug registration and approval within the European Union (EU) has shifted away from the national licensing authorities of EU members to the centralized authority of the European Medicines Evaluation Agency (EMEA), which was established in 1993 [172]. To facilitate this centralized approval process, the EMEA pushed for harmonization of drug approvals. While the goals of harmonization are to create a single pharmaceutical market in Europe and to shorten approval times, concerns were voiced that harmonized safety standards would lower the stricter standards that were favored by some countries such as Sweden, and would compromise patient safety [173]. Now called the European Medicines Agency (EMA), this is a decentralized EU body responsible for the scientific evaluation and supervision of medicines. These functions are performed by the EMA's Committee for Medicinal Products for Human Use (CHMP). EMA authorization to market a drug is valid in all EU countries, but individual national medicines agencies are responsible for monitoring the safety of approved drugs and sharing this information with the EMA [174].

Potential Contributions of Pharmacoepidemiology

The potential contributions of pharmacoepidemiology are now well recognized, even though the field is still relatively new. However, some contributions are already apparent (see Table 1.1). In fact, in the 1970s the FDA requested postmarketing research at the time of approval for about one third of drugs, compared to over 70% in the 1990s [175]. Since the passage

Table 1.1 Potential contributions of pharmacoepidemiology.

- A) Information which supplements the information available from premarketing studies – better quantitation of the incidence of known adverse and beneficial effects
  - 1) Higher precision
  - 2) In patients not studied prior to marketing, e.g., the elderly, children, pregnant women
  - 3) As modified by other drugs and other illnesses
  - 4) Relative to other drugs used for the same indication
- B) New types of information not available from premarketing studies
  - 1) Discovery of previously undetected adverse and beneficial effects
    - i) Uncommon effects
    - ii) Delayed effects
  - 2) Patterns of drug utilization
  - 3) The effects of drug overdoses
  - 4) The economic implications of drug use
- C) General contributions of pharmacoepidemiology
  - 1) Reassurances about drug safety
  - 2) Fulfillment of ethical and legal obligations

of PDUFA IV, the FDA has the right to require that such studies be completed. In this section of this chapter, we will first review the potential for pharmacoepidemiologic studies to supplement the information available prior to marketing, and then review the new types of information obtainable from postmarketing pharmacoepidemiologic studies, but not obtainable prior to drug marketing. Finally, we will review the general, and probably most important, potential contributions such studies can make. In each case, the relevant information available from premarketing studies will be briefly examined first, to clarify how postmarketing studies can supplement it.

Supplementary Information

Premarketing studies of drug effects are necessarily limited in size. After marketing, nonexperimental epidemiologic studies can be performed, evaluating the effects of drugs administered as part of ongoing medical care. These allow the cost-effective accumulation of much larger numbers of patients than those studied prior to marketing, resulting in a more precise measurement of the incidence of adverse and beneficial drug effects (see Chapter 4). For example, at the time of drug marketing, prazosin was known to cause a dose-dependent first dose syncope [176,177], but the FDA requested that the manufacturer conduct a postmarketing surveillance study of the drug in the US to quantitate its incidence more precisely [122]. In recent years, there has even been an attempt, in selected special cases, to release critically important drugs more quickly by taking advantage of the work that can be performed after marketing. Probably the best-known early example was zidovudine [178,179]. More recently, this has been the case with a number of cancer drugs, including at least one where initial expectations of efficacy were not confirmed in definitive trials after marketing, and were then proven again later in a subgroup, leading to the product being removed from the market and then marketed again. As already noted, the increased sample size available after marketing also permits a more precise determination of the correct dose to be used [99,101,180,181]. The study of drug interactions, as previously discussed, is analogous (see also Chapter 40).

Premarketing studies also tend to be very artificial. Important subgroups of patients are not typically included in studies conducted before drug marketing, usually for ethical reasons. Examples include the elderly, children, and pregnant women. Studies of the effects of drugs in these populations generally must be conducted after drug marketing [182]. (See also Chapter 22.)

Additionally, for reasons of statistical efficiency, premarketing clinical trials generally seek subjects who are as homogenous as possible, in order to reduce unexplained variability in the outcome variables measured and increase

the probability of detecting a difference between the study groups, if one truly exists. For these reasons, certain patients are often excluded, including those with other illnesses or those who are receiving other drugs. Postmarketing studies can explore how factors such as other illnesses and other drugs might modify the effects of the drugs, as well as looking at the effects of differences in drug regimen, adherence, and so on [183]. For example, after marketing, the ophthalmic preparation of timolol was noted to cause many serious episodes of heart block and asthma, resulting in more than 10 deaths. These effects were not detected prior to marketing, as patients with underlying cardiovascular or respiratory disease were excluded from the premarketing studies [184].

Finally, to obtain approval to market a drug, a manufacturer needs to evaluate its overall safety and efficacy, but does not need to evaluate its safety and efficacy relative to any other drugs available for the same indication. To the contrary, with the exception of illnesses that could not ethically be treated with placebos, such as serious infections and malignancies, it is generally considered preferable, or even mandatory, to have studies with placebo controls. There are a number of reasons for this preference. First, it is easier to show that a new drug is more effective than a placebo than to show that it is more effective than another effective drug. Second, one cannot actually prove that a new drug is as effective as a standard drug. A study showing that a new drug is no worse than another effective drug does not provide assurance that it is better than a placebo; one simply could have failed to detect that it was in fact worse than the standard drug. One could require a demonstration that a new drug is more effective than another effective drug, but this is a standard that does not and should not have to be met. Yet, optimal medical care requires information on the effects of a drug relative to the alternatives available for the same indication. This information must often await studies conducted after drug marketing. Indeed, as noted, this is a major component of the new focus on CER (see Chapter 26).

New Types of Information Not Available from Premarketing Studies

As already mentioned, premarketing studies are necessarily limited in size (see also Chapter 4). The additional sample size available in postmarketing studies permits the study of drug effects that may be uncommon but important, such as drug-induced agranulocytosis [185].

Premarketing studies are also necessarily limited in time; they must come to an end, or the drug could never be marketed. In contrast, postmarketing studies permit the study of delayed drug effects, such as the unusual clear cell adenocarcinoma of the vagina and cervix, which occurred two decades later in women exposed *in utero* to diethylstilbestrol [15].

The patterns of physician prescribing and patient drug utilization often cannot be predicted prior to marketing, despite pharmaceutical manufacturers' best attempts to predict when planning for drug marketing. Studies of how a drug is actually being used, and determinants of changes in these usage patterns, can only be performed after drug marketing (see Chapters 18 and 19).

In most cases, premarketing studies are performed using selected patients who are closely observed. Rarely are there any significant overdoses in this population. Thus, the study of the effects of a drug when ingested in extremely high doses is rarely possible before drug marketing. Again, this must await postmarketing pharmacoepidemiologic studies [186].

Finally, it is only in the past decade or two that pharmacoepidemiologists have become more sensitive to the costs of medical care, and the techniques of health economics been applied to evaluate the cost implications of drug use [187]. It is clear that exploration of the costs of drug use requires consideration of more than just the

costs of the drugs themselves. The costs of a drug's adverse effects may be substantially higher than the cost of the drug itself, if these adverse effects result in additional medical care and possibly even hospitalizations [188]. Conversely, a drug's beneficial effects could reduce the need for medical care, resulting in savings that could be much larger than the cost of the drug itself. As with studies of drug utilization, the economic implications of drug use can be predicted prior to marketing, but can only be rigorously studied after marketing (see Chapter 34).

General Contributions of Pharmacoepidemiology

Lastly, it is important to review the general contributions that pharmacoepidemiology can make. As an academic or a clinician, one is most interested in the new information about drug effects and drug costs that can be gained from pharmacoepidemiology. Certainly, these are the findings that receive the greatest public and political attention. However, often no new information is obtained, particularly about new adverse drug effects. This is not a disappointing outcome, but in fact a very reassuring one, and this reassurance about drug safety is one of the most important contributions that pharmacoepidemiologic studies can make. Related to this is the reassurance that the sponsor of the study, whether manufacturer or regulator, is fulfilling its organizational duty ethically and responsibly by looking for any undiscovered problems that may exist. In an era of product liability litigation, this is an important assurance. One cannot change whether a drug causes an adverse reaction, and the fact that it does will hopefully eventually become evident. What can be changed is the perception about whether a manufacturer did everything possible to detect it and was not negligent in its behavior.

References

1 Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 1998; **279**: 1200–5.

2 Moore RD, Smith CR, Lietman PS. Increased risk of renal dysfunction due to interaction of liver disease and aminoglycosides. *Am J Med* 1986; **80**: 1093–7.

3 Rawlins MD, Thompson JW. Pathogenesis of adverse drug reactions. In: Davies DM, ed. *Textbook of Adverse Drug Reactions*. Oxford: Oxford University Press, 1977, p. 44.

4 Strom BL. Integrating pharmacoepidemiology into the design and conduct of clinical trials. *J Clin Res Drug Dev* 1988; **2**: 161–8.

5 Geiling EMK, Cannon PR. Pathogenic effects of elixir of sulfanilimide (diethylene glycol) poisoning. *JAMA* 1938; **111**: 919–26.

6 Wallerstein RO, Condit PK, Kasper CK, Brown JW, Morrison FR. Statewide study of chloramphenicol therapy and fatal aplastic anemia. *JAMA* 1969; **208**: 2045–50.

7 Meyler L. *Side Effects of Drugs*. Amsterdam: Elsevier, 1952.

8 Erslev AJ, Wintrobe MM. Detection and prevention of drug induced blood dyscrasias. *JAMA* 1962; **181**: 114–19.

9 Cluff LE, Thornton GF, Seidl LG. Studies on the epidemiology of adverse drug reactions. I. Methods of surveillance. *JAMA* 1964; **188**: 976–83.

10 Miller RR, Greenblatt DJ. *Drug Effects in Hospitalized Patients*. New York: John Wiley & Sons, Inc, 1976.

11 Caranasos GJ, Stewart RB, Cluff LE. Drug-induced illness leading to hospitalization. *JAMA* 1974; **228**: 713–17.

12 Lenz W. Malformations caused by drugs in pregnancy. *Am J Dis Child* 1966; **112**: 99–106.

13 Wardell WM. Therapeutic implications of the drug lag. *Clin Pharmacol Ther* 1974; **15**: 73–96.

14 Kono R. Trends and lessons of SMON research. In: Soda T, ed. *Drug-Induced Sufferings*. Princeton, NJ: Excerpta Medica, 1980, p. 11.

15 Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971; **284**: 878–81.

16 Wright P. Untoward effects associated with practolol administration. Oculomucocutaneous syndrome. *BMJ* 1975; **1**: 595–8.

17 Ticrynafen recalled. *FDA Drug Bull* 1980; **10**: 3–4.

18 Suspension of benoxaprofen (Opren). *BMJ* 1982; **285**: 519–20.

19 Strom BL, Carson JL, Morse ML, West SL, Soper KA. The effect of indication on hypersensitivity reactions associated with zomepirac sodium and other nonsteroidal antiinflammatory drugs. *Arthritis Rheum* 1987; **30**: 1142–8.

20 Inman WHW. Study of fatal bone marrow depression with special reference to phenylbutazone and oxyphenbutazone. *BMJ* 1977; **1**: 1500–5.

21 Committee on Safety of Medicines. *Osmosin (Controlled release indomethacin). Current Problems*, no. 11. London, 1983.

22 Barash CI, Lasagna L. The Bendectin saga: "Voluntary" discontinuation. *J Clin Res Drug Devel* 1987; **1**: 277–92.

23 Strom BL, West SL, Sim E, Carson JL. The epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693–9.

24 FDA ponders approaches to curbing adverse effects of drug used against cystic acne. *JAMA* 1988; **259**: 3225–30.

25 Stern RS. When a uniquely effective drug is teratogenic: the case of isotretinoin. *N Engl J Med* 1989; **320**: 1007–9.

26 Hertzman PA, Falk H, Kilbourne EM, Page S, Shulman LE. The Eosinophilia-Myalgia syndrome: the Los Alamos Conference. *J Rheumatol* 1991; **18**: 867–73.

27 Meyboom RHB. The triazolam experience in 1979 in The Netherlands, a problem of signal generation and verification. In: Strom BL, Velo GP, eds. *Drug Epidemiology and Postmarketing Drug Surveillance*. New York: Plenum Press, 1992, pp. 159–67.

28 Triazolam's status in European Community. *Lancet* 1991; **338**: 1586–7.

29 Brahams D. Triazolam suspended. *Lancet* 1991; **338**: 938.

30 Triazolam hearing. *Lancet* 1992; **339**: 1291.

31 Randall T. Penile, testicular, other silicone implants soon will undergo FDA review (news). *JAMA* 1992; **267**: 2578–9.

32 Egger M, Smith GD, Imhoof H, Teuscher A. Risk of severe hypoglycaemia in insulin treated diabetic patients transferred to human insulin: a case control study. *BMJ* 1991; **303**: 617–21.

33 Egger M, Smith GD, Teuscher AU, Teuscher A. Influence of human insulin on symptoms and awareness of hypoglycaemia: a randomised double blind crossover trial. *BMJ* 1991; **303**: 622–6.

34 Colagiuri S, Miller JJ, Petocz P. Double-blind crossover comparison of human and porcine insulins in patients reporting lack of hypoglycaemia awareness. *Lancet* 1992; **339**: 1432–5.

35 Jick H, Hall GC, Dean AD, Jick SS, Derby LE. A comparison of the risk of hypoglycemia between users of human and animal insulins. 1. Experience in the United Kingdom. *Pharmacotherapy* 1990; **10**: 395–7.

36 Jick SS, Derby LE, Gross KM, Jick H. Hospitalizations because of hypoglycemia in users of animal and human insulins. 2.

Experience in the United States. *Pharmacotherapy* 1990; **10**: 398–9.

37 Teicher MH, Glod C, Cole JO. Emergence of intense suicidal preoccupation during fluoxetine treatment. *Am J Psychiatry* 1990; **147**: 207–10.

38 Beasley CM, Dornseif BE, Bosomworth JC, *et al*. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *BMJ* 1991; **303**: 685–92.

39 Crane J, Pearce N, Flatt A, *et al*. Prescribed fenoterol and death from asthma in New Zealand, 1981–83: case-control study. *Lancet* 1989; **1**: 917–22.

40 Pearce N, Grainger J, Atkinson M, *et al*. Case-control study of prescribed fenoterol and death from asthma in New Zealand, 1977–81. *Thorax* 1990; **45**: 170–5.

41 Grainger J, Woodman K, Pearce N, *et al*. Prescribed fenoterol and death from asthma in New Zealand, 1981–7: a further case-control study. *Thorax* 1991; **46**: 105–11.

42 Spitzer WO, Suissa S, Ernst P, *et al*. The use of beta-agonists and the risk of death and near death from asthma. *N Engl J Med* 1992; **326**: 501–6.

43 Rosenfield A, Maine D, Rochat R, Shelton J, Hatcher RA. The Food and Drug Administration and medroxyprogesterone acetate. What are the issues? *JAMA* 1983; **249**: 2922–8.

44 WHO Collaborative Study of Neoplasia and Steroid Contraceptives. Breast cancer and depot-medroxyprogesterone acetate: a multinational study. *Lancet* 1991; **338**: 833–8.

45 WHO Collaborative Study of Neoplasia and Steroid Contraceptives. Depot-medroxyprogesterone acetate (DMPA) and risk of invasive squamous cell cervical cancer. *Contraception* 1992; **45**: 299–312.

46 Nightingale SL. From the Food and Drug Administration: warnings issued on nonsedating antihistamines terfenadine and astemizole. *JAMA* 1992; **268**: 705.

47 Ahmad SR. Antihistamines alert. *Lancet* 1992; **340**: 542.

48 Rothman KJ, Funch DP, Dreyer NA. Bromocriptine and puerperal seizures. *Epidemiology* 1990; **1**: 232–8.

49 Gross TP. Bromocriptine and puerperal seizures. *Epidemiology* 1991; **2**: 234–5.

50 Finch RG. The withdrawal of temafloxacin: are there implications for other quinolones? *Drug Saf* 1993; **8**: 9–11.

51 Nathani MG, Mutchnick MG, Tynes DJ, Ehrinpreis MN. An unusual case of amoxicillin/clavulanic acid-related hepatotoxicity. *Am J Gastroenterol* 1998; **93**: 1363–5.

52 Hunter EB, Johnston PE, Tanner G, Pinson CW, Awad JA. Bromfenac (Duract)-associated hepatic failure requiring liver transplantation. *Am J Gastroenterol* 1999; **94**: 2299–301.

53 Moses PL, Schroeder B, Alkhatib O, Ferrentino N, Suppan T, Lidofsky SD. Severe hepatotoxicity associated with bromfenac sodium. *Am J Gastroenterol* 1999; **94**: 1393–6.

54 Furberg CD, Psaty BM, Meyer JV. Nifedipine: dose-related increase in mortality in patients with coronary heart disease. *Circulation* 1995; **92**: 1326–31.

55 Pahor M, Guralnik JM, Ferrucci L, *et al*. Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348**: 493–7.

56 Braun S, Boyko V, Behar S, *et al*. Calcium antagonists and mortality in patients with coronary artery disease: a cohort study of 11,575 patients. *J Am Coll Cardiol* 1996; **28**: 7–11.

57 Kostis JB, Lacy CR, Cosgrove NM, Wilson AC. Association of calcium channel blocker use with increased rate of acute myocardial infarction in patients with left ventricular dysfunction. *Am Heart J* 1997; **133**: 550–7.

58 Jick H, Jick S, Derby LE, Vasilakis C, Myers MW, Meier CR. Calcium-channel blockers and risk of cancer. *Lancet* 1997; **349**: 525–8.

59 Fitzpatrick AL, Daling JR, Furberg CD, Kronmal RA, Weissfeld JL. Use of calcium channel blockers and breast carcinoma risk in postmenopausal women. *Cancer* 1997; **80**: 1438–47.

60 Rosenberg L, Rao RS, Palmer JR, *et al*. Calcium channel blockers and the risk of cancer. *JAMA* 1998; **279**: 1000–4.

61 Michels KB, Rosner BA, Walker AM, *et al*. Calcium channel blockers, cancer incidence, and cancer mortality in a cohort of U.S. women: the Nurses' Health Study. *Cancer* 1998; **83**: 2003–7.

62 Wysowski DK, Bacsanyi J. Cisapride and fatal arrhythmia. *N Engl J Med* 1996; **335**: 290–1.

63 Piquette RK. Torsade de pointes induced by cisapride/clarithromycin interaction. *Ann Pharmacother* 1999; **33**: 22–6.

64 Smalley W, Shatin D, Wysowski DK, *et al*. Contraindicated use of cisapride: impact of Food and Drug Administration regulatory action. *JAMA* 2000; **284**: 3036–9.

65 Weatherby LB, Walker AM, Fife D, Vervaet P, Klausner MA. Contraindicated medications dispensed with cisapride: temporal trends in relation to the sending of "Dear Doctor" letters. *Pharmacoepidemiol Drug Saf* 2001; **10**: 211–18.

66 Kramer MS, Lane DA. Aminorex, dexfenfluramine, and primary pulmonary hypertension. *J Clin Epidemiol* 1998; **51**: 361–4.

67 Connolly HM, Crary JL, McGoon MD, *et al*. Valvular heart disease associated with fenfluramine-phentermine. *N Engl J Med* 1997; **337**: 581–8.

68 Kimmel SE, Keane MG, Crary JL, *et al*. Detailed examination of fenfluramine-phentermine users with valve abnormalities identified in Fargo, North Dakota. *Am J Cardiol* 1999; **84**: 304–8.

69 Strom BL, Berlin JA, Kinman JL, *et al*. Parenteral ketorolac and risk of gastrointestinal and operative site bleeding: a postmarketing surveillance study. *JAMA* 1996; **275**: 376–82.

70 Feldman HI, Kinman JL, Berlin JA, *et al*. Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193–9.

71 Hennessy S, Kinman JL, Berlin JA, *et al*. Lack of hepatotoxic effects of parenteral ketorolac in the hospital setting. *Arch Intern Med* 1997; **157**: 2510–14.

72 Garcia Rodriguez LA, Cattaruzzi C, Troncon MG, Agostinis L. Risk of hospitalization for upper gastrointestinal tract bleeding associated with ketorolac, other nonsteroidal anti-inflammatory drugs, calcium antagonists, and other antihypertensive drugs. *Arch Intern Med* 1998; **158**: 33–9.

73 Mullins ME, Horowitz BZ, Linden DH, Smith GW, Norton RL, Stump J. Life-threatening interaction of mibefradil and beta-blockers with dihydropyridine calcium channel blockers. *JAMA* 1998; **280**: 157–8.

74 Spitzer WO, Lewis MA, Heinemann LA, Thorogood M, MacRae KD. Third generation oral contraceptives and risk of venous thromboembolic disorders: an international case-control study. *BMJ* 1996; **312**: 83–8.

75 Jick H, Jick SS, Gurewich V, Myers MW, Vasilakis C. Risk of idiopathic cardiovascular death and nonfatal venous thromboembolism in women using oral contraceptives with differing progestagen components. *Lancet* 1995; **346**: 1589–93.

76 WHO Technical Report Series. *Cardiovascular Disease and Steroid Hormone Contraception*. Geneva: World Health Organization, 1998.

77 de Bruijn SF, Stam J, Vandenbroucke JP. Increased risk of cerebral venous sinus thrombosis with third-generation oral contraceptives. *Lancet* 1998; **351**: 1404.

78 Feenstra J, van Drie-Pierik RJ, Lacle CF, Stricker BH. Acute myocardial infarction associated with sildenafil. *Lancet* 1998; **352**: 957–8.

79 Kahn LH, Alderfer RJ, Graham DJ. Seizures reported with tramadol. *JAMA* 1997; **278**: 1661.

80 Jick H, Derby LE, Vasilakis C, Fife D. The risk of seizures associated with tramadol. *Pharmacotherapy* 1998; **18**: 607–11.

81 Pereira SP, Williams R. Adverse events associated with Vitamin K_1: results of a worldwide postmarketing surveillance programme. *Pharmacoepidemiol Drug Saf* 1998; **7**: 173–82.

82 Misbin RI. Troglitazone-associated hepatic failure. *Ann Intern Med* 1999; **130**: 330.

83 Vella A, de Groen PC, Dinneen SF. Fatal hepatotoxicity associated with troglitazone. *Ann Intern Med* 1998; **129**: 1080.

84 Neuschwander-Tetri BA, Isley WL, Oki JC, *et al*. Troglitazone-induced hepatic failure leading to liver transplantation: a case report. *Ann Intern Med* 1998; **129**: 38–41.

85 Gitlin N, Julie NL, Spurr CL, Lim KN, Juarbe HM. Two cases of severe clinical and histologic hepatotoxicity associated with troglitazone. *Ann Intern Med* 1998; **129**: 36–8.

86 Centers for Disease Control and Prevention. Suspension of rotavirus vaccine after reports of intussusception—United States, 1999. *Morb Mortal Wkly Rep* 2004; **53**: 786–9.

87 Moynihan R. Alosetron: a case study in regulatory capture, or a victory for patients' rights? *BMJ* 2002; **325**: 592–5.

88 Thompson PD, Clarkson P, Karas RH. Statin-associated myopathy. *JAMA* 2003; **289**: 1681–90.

89 Schulman SR. Rapacuronium redux. *Anesth Analg* 2002; **94**: 483–4.

90 Kelly DL, Love RC. Ziprasidone and the QTc interval: pharmacokinetic and pharmacodynamic considerations. *Psychopharmacol Bull* 2001; **35**: 66–79.

91 Kernan WN, Viscoli CM, Brass LM, *et al*. Phenylpropanolamine and the risk of hemorrhagic stroke. *N Engl J Med* 2000; **343**: 1826–32.

92 Lathrop SL, Ball R, Haber P, *et al*. Adverse event reports following vaccination for Lyme disease: December 1998–July 2000. *Vaccine* 2002; **20**: 1603–8.

93 Joellenbeck LM, Zwanziger LL, Durch JS, Strom BL, eds.; for Committee to Assess the Safety and Efficacy of the Anthrax Vaccine. *The Anthrax Vaccine: Is It safe? Does It Work?* Washington, DC: National Academies Press, 2002.

94 Chen RT, Lane JM. Myocarditis: the unexpected return of smallpox vaccine adverse events. *Lancet* 2003; **362**: 1345–6.

95 Topol EJ. Failing the public health—rofecoxib, Merck, and the FDA. *N Engl J Med* 2004; **351**: 1707–9.

96 Shah RR. The significance of QT interval in drug development. *Br J Clin Pharmacol* 2002; **54**: 188–202.

97 U.S. Food & Drug Administration. Drug Safety-Related Labeling Changes. http://www.fda.gov/Safety/MedWatch/SafetyInformation/Safety-RelatedDrugLabelingChanges/default.htm (accessed April 2019).

98 United States Government Accountability Office. *FDA Drug Review: Postapproval Risks 1976–1985*. GAO/PEMD-90-15, April 1990, p. 24.

99 Cross J, Lee H, Westelinck A, Nelson J, Grudzinskas C, Peck C. Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999. *Pharmacoepidemiol Drug Saf* 2002; **11**: 439–46.

100 United States Government Accountability Office. Drug Safety: Improvement Needed in FDA's Postmarket Decision-Making and Oversight Process. GAO-06-402. March 2006. http://www.gao.gov/new.items/d06402.pdf (accessed April 2019).

101 Heerdink ER, Urquhart J, Hubert G, Leufkens HG. Changes in prescribed drug doses after market introduction. *Pharmacoepidemiol Drug Saf* 2002; **11**: 447–453.

102 Wardell WM. Therapeutic implications of the drug lag. *Clin Pharmacol Ther* 1974; **15**(1): 73–96.

103 US Food & Drug Administration. *Prescription Drug User Fee Act (PDUFA)*. http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/default.htm (accessed November 2010).

104 Berndt ER, Gottschalk AH, Philipson TJ, Strobeck MW. Industry funding of the FDA: effects of PDUFA on approval times and withdrawal rates. *Nat Rev Drug Discov* 2005; **4**(7): 545–54.

105 Friedman MA, Woodcock J, Lumpkin MM, Shuren JE, Hass AE, Thompson LJ. The safety of newly approved medicines: do recent market removals mean there is a problem? *JAMA* 1999; **281**: 1728–34.

106 Kleinke JD, Gottlieb S. Is the FDA approving drugs too fast? Probably not—but drug recalls have sparked debate. *BMJ* 1998; **317**: 899.

107 Lurie P, Woodcock J, Kaitin KI. FDA drug review: the debate over safety, efficacy, and speed. *Medical Crossfire* 1999; **1**: 52–60.

108 Wood AJ, Stein CM, Woosley R. Making medicines safer—the need for an independent drug safety board. *N Engl J Med* 1998; **339**: 1851–4.

109 Moore TJ, Psaty BM, Furberg CD. Time to act on drug safety. *JAMA* 1998; **279**: 1571–3.

110 Institute of Medicine Committee on the Assessment of the US Drug Safety System. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC: National Academy of Sciences, 2007.

111 FDAnews. PDUFA IV fees will go to improve postmarket drug safety. *FDAnews Drug Daily Bulletin* 2008; 5. http://www.fdanews.com/newsletter/article?articleId=106457&issueId=11559 (accessed April 2019).

112 U.S. Food & Drug Administration. *Food and Drug Administration Safety and Innovation Act (FDASIA)*. https://www.fda.gov/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDASIA/default.htm (accessed April 2019).

113 U.S. Food & Drug Administration. *Generic Drug User Fee Amendments*. https://www.fda.gov/ForIndustry/UserFees/GenericDrugUserFees/default.htm (accessed April 2019).

114 U.S. Food & Drug Admniistration. *Biosimilar User Fee Act (BsUFA)*. https://www.fda.gov/ForIndustry/UserFees/BiosimilarUserFeeActBsUFA/default.htm (accessed April 2019).

115 Congress.gov. H.R.6 – *21st Century Cures Act*. https://www.congress.gov/bill/114th-congress/house-bill/6 (accessed April 2019).

116 Lee JAH, Draper PA, Weatherall M. Prescribing in three English towns. *Milbank Mem Fund Quart* 1965; **43**: 285–90.

117 Muller C. Medical review of prescribing. *J Chronic Dis* 1965; **18**: 689–96.

118 Meade TW. Prescribing of chloramphenicol in general practice. *BMJ* 1967; **1**: 671–4.

119 Joyce CRB, Last JM, Weatherall M. Personal factors as a cause of differences in prescribing by general practitioners. *Br J Prev Soc Med* 1968; **22**: 170–7.

120 Muller C. Outpatient drug prescribing related to clinic utilization in four New York City hospitals. *Health Serv Res* 1968; **3**: 142–54.

121 Slone D, Shapiro S, Miettinen OS. Case-control surveillance of serious illnesses attributable to ambulatory drug use. In: Colombo F, Shapiro S, Slone D, Tognoni G, eds. *Epidemiological Evaluation of Drugs*. Littleton, MA: PSG, 1977, pp. 59–82.

122 Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC: U.S. Government Printing Office, 1980.

123 Strom BL, Carson JL, Morse ML, Leroy AA. The Computerized Online Medicaid Analysis and Surveillance System: a new resource for post-marketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.

124 Inman WHW. Prescription event monitoring. *Acta Med Scand Suppl* 1984; **683**: 119–26.

125 Strom BL, Members of the ASCPT Pharmacoepidemiology Section. Position paper on the use of purported postmarketing drug surveillance studies for promotional purposes. *Clin Pharmacol Ther* 1990; **48**: 598.

126 Andrews EA, Avorn J, Bortnichak EA, *et al.*; ISPE. Guidelines for good epidemiology practices for drug, device, and vaccine research in the United States. *Pharmacoepidemiol Drug Saf* 1996; **5**: 333–8.

127 ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.

128 Neutel CI. Privacy issues in research using record linkage. *Pharmacoepidemiol Drug Saf* 1997; **6**: 367 –9.

129 Melton LJ III. The threat to medical-records research. *N Engl J Med* 1997; **337**: 1466–70.

130 Mann RD. Data privacy and confidentiality. *Pharmacoepidemiol Drug Saf* 1999; **8**: 245.

131 Andrews EB. Data privacy, medical record confidentiality, and research in the interest of public health. *Pharmacoepidemiol Drug Saf* 1999; **8**: 247–60.

132 Mann RD. The issue of data privacy and confidentiality in Europe – 1998. *Pharmacoepidemiol Drug Saf* 1999; **8**: 261–4.

133 Califf RM. The need for a national infrastructure to improve the rational use of therapeutics. *Pharmacoepidemiol Drug Saf* 2002; **11**: 319–27.

134 Congress.gov. *H.R.1 – Medicare Prescription Drug, Improvement, and Modernization Act of 2003. Sec. 1013. Research on Outcomes of Health Care Items and Services*. https://www.congress.gov/bill/108th-congress/house-bill/00001 (accessed April 2019).

135 Agency for Healthcare Research and Quality. The Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) research network (U19). 2013. https://grants.nih.gov/grants/guide/rfa-files/RFA-HS-13-009.html (accessed April 2019).

136 Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academies Press, 2000.

137 Bates DW, Cullen DJ, Laird N, *et al.* Incidence of adverse drug events and potential adverse drug events: implications for prevention. *JAMA* 1995; **274**: 29–34.

138 Brennan TA, Leape LL, Laird NM, *et al.* Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *N Engl J Med* 1991; **324**: 370–6.

139 Classen DC, Pestotnik SL, Evans R, Scott R, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients: excess length of stay, extra costs, and attributable mortality. *JAMA* 1997; **277**: 301–6.

140 Bates DW, Spell N, Cullen DJ, *et al.* The costs of adverse drug events in hospitalized patients. *JAMA* 1997; **277**: 307–11.

141 Phillips DP, Christenfeld N, Glynn LM. Increase in US medication-error deaths between 1983 and 1993. *Lancet* 1998; **351**: 643–4.

142 Manasse HR. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1655.

143 Ferner RE, Anton C. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1655–6.

144 Rooney C. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1656–7.

145 Phillips DP, Christenfeld N, Glynn LM. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1657.

146 McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in the Institute of Medicine report. *JAMA* 2000; **284**: 93–5.

147 Leape LL. Institute of Medicine error figures are not exaggerated. *JAMA* 2000; **284**: 95–7.

148 Manson JE, Hsia J, Johnson KC, *et al.* Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; **349**: 523–34.

149 Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; **288**: 321–33.

150 Col NF, Pauker SG. The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? *Ann Intern Med* 2003; **139**: 923–9.

151 Grimes DA, Lobo RA. Perspectives on the Women's Health Initiative trial of hormone replacement therapy. *Obstet Gynecol* 2002; **100**: 1344–53.

152 Whittemore AS, McGuire V. Observational studies and randomized trials of hormone replacement therapy: what can we learn from them? *Epidemiology* 2003; **14**: 8–9.

153 Piantadosi S. Larger lessons from the Women's Health Initiative. *Epidemiology* 2003; **14**: 6–7.

154 United States Government Accountability Office. New Drug Approval: FDA needs to enhance its oversight of drugs approved on the basis of surrogate endpoints. *GAO Highlights*, September 23, 2009. http://www.gao.gov/highlights/d09866high.pdf (accessed April 2019).

155 Federal Coordinating Council for Comparative Effectiveness Research. *Report to the President and Congress on Comparative Effectiveness Research*. Washington, DC: US Department of Health and Human Services, 2009.

156 National Institute for Health and Clinical Excellence. What We Do. http://www.nice.org.uk/aboutnice/whatwedo/what_we_do.jsp (accessed April 2019).

157 Reichert JM. Trends in development and approval times for new therapeutics in the United States. *Nat Rev Drug Discov* 2003; **2**(9): 695–702.

158 Haberman AB. Overcoming Phase II attrition problem. *Genet Eng Biotechn N* 2009; **29**: 1.

159 Pharmaceutical Research and Manufacturers of America (PhRMA). *Pharmaceutical Industry Profile 2003*. Washington, DC: PhRMA, 2003, p. 75.

160 Pharmaceutical Research and Manufacturers of America (PhRMA). *Pharmaceutical Industry Profile 2009*. Washington, DC: PhRMA, 2003, p. 2.

161 Shafrin J. Cost to bring drug to market: $802m. *Healthcare Economist*, April 29, 2006. http://healthcare-economist.com/2006/04/29/802m (accessed April 2019).

162 Collier R. Drug development cost estimates hard to swallow. *CMAJ* 2009; **180**: 279–80.

163 Ratanawijitrasin S, Wondemagegnehu E. *Effective Drug Regulation: A Multicountry Study*. Geneva: World Health Organization, 2002. http://archives.who.int/tbs/qual/s2300e.pdf (accessed April 2019).

164 Government of Canada. Health Canada. http://www.hc-sc.gc.ca/index-eng.php (accessed April 2019).

165 China Food and Drug Administration. https://en.wikipedia.org/wiki/China_Food_and_Drug_Administration (accessed April 2019)

166 Danish Medicines Agency. *About Us*. http://www.dkma.dk/1024/visUKLSForside.asp?artikelID=728 (accessed April 2019).

167 Norwegian Medicines Agency. About Us. http://www.legemiddelverket.net/templates/InterPage____16645.aspx (accessed April 2019).

168 Pacific Bridge Medical. *Singapore's Pharmaceutical and Medical Device Regulatory Environment: 2005 Update*. https://www.pacificbridgemedical.com/publication/singapore-8217-s-pharmaceutical-and-medical-device-regulatory-environment-2005-update (accessed April 2019).

169 New Zealand Medicines and Medical Devices Safety Authority. About Medsafe. http://www.medsafe.govt.nz/other/about.asp (accessed April 2019).

170 Central Drugs Standard Control Organization. About CDSCO. https://cdsco.gov.in/opencms/opencms/en/Home (accessed April 2019).

171 Reichert JM. Pharmacopolitics: drug regulation in the United States and Germany. *N Engl J Med* 2004; **350**: 2220–1.

172 European Medicines Agency. *About Us*. https://www.ema.europa.eu/en/about-us (accessed April 2019).

173 Abraham J, Lewis G. Harmonising and competing for medicines regulation: how healthy are the European Union's systems of drug approval? *Soc Sci Med* 1999; **48**: 1655–67.

174 European Medicines Agency. *What We Do*. https://www.ema.europa.eu/en/about-us/what-we-do (accessed April 2019).

175 Department of Health and Human Services, Office of Inspector General. *Postmarketing Studies of Prescription Drugs*. OEI-03-94-00760. Philadelphia, PA: OIG. May 1996, p. 11. http://oig.hhs.gov/oei/reports/oei-03-94-00760.pdf (accessed April 2019).

176 Rosendorff C. Prazosin: severe side effects are dose-dependent. *BMJ* 1976; **2**: 508.

177 Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first-dose phenomenon. *BMJ* 1976; **2**: 1293–4.

178 Kaitin KI. Case studies of expedited review: AZT and L-Dopa. *Law Med Health Care* 1991; **19**: 242–6.

179 Wastila LJ, Lasagna L. The history of zidovudine (AZT). *J Clin Res Pharmacoepidemiol* 1990; **4**: 25–37.

180 Peck CC. Postmarketing drug dosage changes. *Pharmacoepidemiol Drug Saf* 2003; **12**: 425–6.

181 Temple RJ. Defining dose decrease. *Pharmacoepidemiol Drug Saf* 2003; **12**: 151–2.

182 McKenzie MW, Marchall GL, Netzloff ML, Cluff LE. Adverse drug reactions leading to hospitalization in children. *J Pediatr* 1976; **89**: 487–90.

183 May FE, Stewart RB, Cluff LE. Drug interactions and multiple drug administration. *Clin Pharmacol Ther* 1977; **22**: 322–8.

184 Nelson WL, Fraunfelder FT, Sills JM, Arrowsmith JB, Kuritsky JN. Adverse respiratory and cardiovascular events attributed to timolol ophthalmic solution, 1978–1985. *Am J Ophthalmol* 1986; **102**: 606–11.

185 International Agranulocytosis and Aplastic Anemia Study. Risks of agranulocytosis and aplastic anemia: a first report of their relation to drug use with special reference to analgesics. *JAMA* 1986; **256**: 1749–57.

186 Stewart RB, Forgnone M, May FE, Forbes J, Cluff LE. Epidemiology of acute drug intoxications: patient characteristics, drugs, and medical complications. *Clin Toxicol* 1974; **7**: 513–30.

187 Eisenberg JM. New drugs and clinical economics: analysis of cost-effectiveness analysis in the assessment of pharmaceutical innovations. *Rev Infect Dis* 1984; **6**(Suppl 4): 905–8.

188 Morse ML, Leroy AA, Gaylord TA, Kellenberger T. Reducing drug therapy-induced hospitalization: impact of drug utilization review. *Drug Inf J* 1982; **16**: 199–202.

2

Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiologic Studies

Jeffrey S. Barrett

Bill and Melinda Gates Medical Research Institute, Cambridge, MA, USA

Generally, pharmacology deals with the study of drugs, while clinical pharmacology deals with the study of drugs in humans. More specifically, clinical pharmacology evaluates the characteristics, effects, properties, reactions, and uses of drugs. Of particular interest is their therapeutic value in humans, including their toxicology, safety, pharmacodynamics, and pharmacokinetics. While the foundation of the discipline is underpinned by basic pharmacology (i.e., the study of the interactions that occur between a living organism and exogenous chemicals that alter normal biochemical function), the important emphasis is the application of pharmacologic principles and methods in the care of patients. It has a broad scope, from the discovery of new target molecules and molecular targets to the evaluation of clinical utility in specific populations. Clinical pharmacology bridges the gap between laboratory science and medical practice. Its main objective is to promote the safe and effective use of drugs, maximizing beneficial drug effects while minimizing harmful side effects. It is important that caregivers are skilled in areas of drug information, medication safety, and other aspects of pharmacy practice related to clinical pharmacology. Clinical pharmacology is an important bridging discipline that necessitates knowledge about dose exposure (pharmacokinetics), exposure response (pharmacodynamics), and response outcomes to define the therapeutic window (i.e., the dosage of a medication between the amount that produces the desired or beneficial effect and the amount that produces more adverse effects than desired effects) of a drug in various patient populations. Likewise, clinical pharmacology principles also guide dose modifications in various patient subpopulations (pediatrics, pregnancy, the elderly, and those with organ impairment) and/or dose adjustments for various lifestyle factors (food, time of day, drug interactions).

The discovery and development of new medicines are reliant upon clinical pharmacology research. Scientists in academic, regulatory, and industrial settings participate in this research as part of the overall drug development process. Likewise, the output from clinical pharmacology investigation appears in the drug monograph or package insert of all new medicines, and forms the basis of how drug dosing information is communicated to healthcare providers.

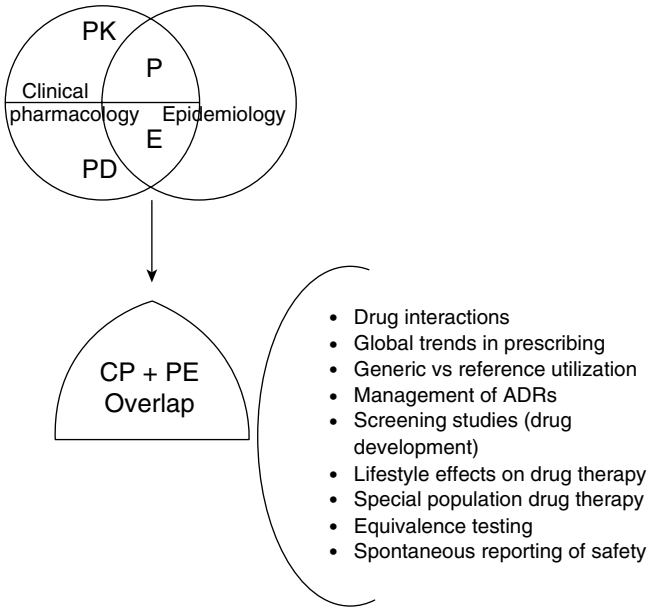
Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Clinical Pharmacology and Pharmacoepidemiology

Pharmacoepidemiology is the study of the utilization and effects of drugs in large numbers of people (see Chapter 1). To accomplish this, pharmacoepidemiology borrows from both clinical pharmacology and epidemiology. Thus, pharmacoepidemiology can also be called a bridging science spanning both clinical pharmacology and epidemiology. Part of the task of clinical pharmacology is to provide a risk–benefit assessment for the effect of drugs in patients. Studies that estimate the probability of beneficial effects in populations, or the probability of adverse effects in populations, will benefit from using epidemiologic methods. Pharmacoepidemiology then can also be defined as the application of epidemiologic methods to the content area of clinical pharmacology. Figure 2.1 illustrates the relationship between clinical pharmacology and pharmacoepidemiology, as well as some of the specific research areas reliant on both disciplines.

Basics of Clinical Pharmacology

Clinical pharmacology encompasses drug composition, drug biopharmaceutic properties, interactions, toxicology, and effects (both desirable and undesirable) that can be used in therapy of diseases. As described earlier, underlying the discipline of clinical pharmacology are the fields of pharmacokinetics and pharmacodynamics, and each of these disciplines can be further defined by the unique processes that dictate composite pathways (e.g., absorption, distribution, metabolism, elimination). Clinical pharmacology is essential both to our understanding of how drugs work as well as how to guide their administration. Individual



Pharmacoepidemiology borrows from both clinical pharmacology and epidemiology. Thus, pharmacoepidemiology can also be called a bridging science spanning both clinical pharmacology and epidemiology. Part of the task of clinical pharmacology is to provide a risk-benefit assessment for the effect of drugs in patients.

Figure 2.1 Relationship between clinical pharmacology and pharmacoepidemiology, illustrating the overlapping areas of interest. ADRs, adverse drug reactions; CP, clinical pharmacology; PD, pharmacodynamics; PE, pharmacoepidemiology; PK, pharmacokinetics.

pharmacotherapy can be challenging due to physiologic factors that may alter drug kinetics (age, size, etc.), pathophysiologic differences that may alter pharmacodynamics, disease subpopulations that may differ from the "mainstream population," and other factors that may result in great variation in safety and efficacy outcomes. The situation becomes even more difficult when one considers critically ill populations and the paucity of well-controlled clinical trials in vulnerable populations. Likewise, healthcare providers who prescribe medications to the critically ill and other difficult-to-manage patients must have some understanding of the basic processes that govern the current dosing recommendations for their patients.

Pharmacokinetics

Pharmacokinetics refers to the study of the mechanisms of absorption and distribution of an administered drug, the chemical changes of the substance in the body (metabolism), and the effects and routes of excretion of the metabolites of the drug (elimination). Each of these subprocesses is defined in greater detail.

Absorption

Absorption is the process of drug transfer from the site of administration to the bloodstream. The rate and efficiency of absorption depend on the route of administration. For intravenous administration, absorption is complete; the total dose reaches the systemic circulation. Drugs administered enterally may be absorbed by either passive diffusion or active transport. The **bioavailability** (F) of a drug is defined by the fraction of the administered dose that reaches the systemic circulation. If a drug is administered intravenously, then the bioavailability is 100% and F = 1.0. When drugs are administered by routes other than intravenously, the bioavailability is usually less. Bioavailability is reduced by incomplete absorption, first-pass metabolism, and distribution into other tissues.

Volume of Distribution

The **volume of distribution** (Vd) is a hypothetical volume of fluid through which a drug is dispersed. A drug rarely disperses solely into the water compartments of the body. Instead, the majority of drugs disperse to several compartments, including adipose tissue and plasma proteins. The total volume into which a drug disperses is called the apparent volume of distribution. This volume is not a physiologic space, but instead a conceptual parameter. It relates the total amount of drug in the body to the concentration of the drug (C) in the blood or plasma: Vd = Drug/C.

Figure 2.2 represents the fate of a drug after intravenous administration. After administration, a maximal plasma concentration is achieved, and the drug is immediately distributed. The plasma concentration then decreases over time. This initial phase is called the alpha (α) phase of drug distribution, where the decline in plasma concentration is due to the distribution of the drug. Once a drug is distributed, it undergoes metabolism and elimination. The second phase is called the beta (β) phase, where the decline in plasma concentration is due to drug metabolism and clearance. The terms A and B are intercepts with the Y axis. The extrapolation of the beta phase defines B. The dotted line is generated by subtracting the extrapolated line from the original concentration line. This second line defines alpha and A. The plasma concentration can be determined using the formula $C = Ae^{-\alpha t} + Be^{-\beta t}$. The distribution and elimination half-lives can be determined by $t_{1/2\alpha} = 0.693/\alpha$ and $t_{1/2\beta} = 0.693/\beta$, respectively [1]. For drugs in which distribution is homogenous along the varied physiologic spaces, the distinction between the alpha and beta phases may be subtle, and essentially a single phase best describes the decline in drug concentration.

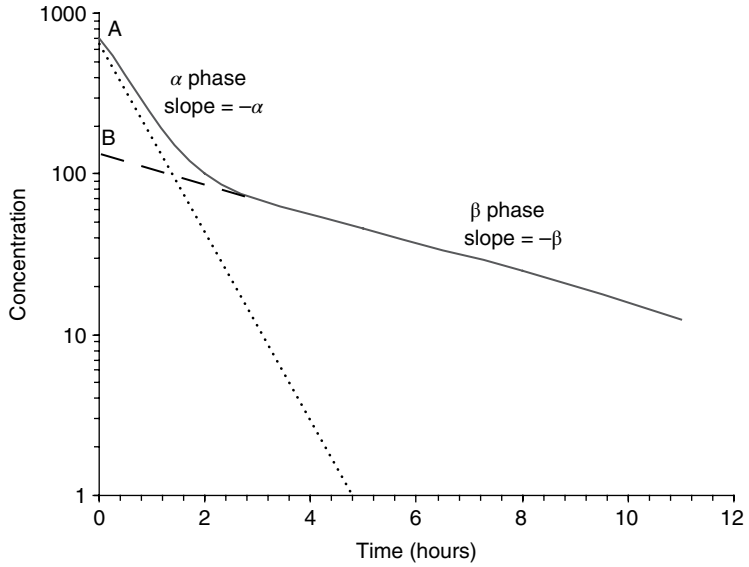


Figure 2.2 Semi-logarithmic plot of concentration vs. time after an intravenous administration of a drug that follows two-compartment pharmacokinetics.

Metabolism

The **metabolism** of drugs is catalyzed by enzymes, and most reactions follow Michaelis Menten kinetics: V(rate of drug metabolism) = [((Vmax)(C)/Km) + (C)], where C is the drug concentration and Km is the Michaelis Menten constant [1]. In most situations, the drug concentration is much less than Km and the equation simplifies to V = (Vmax)(C)/Km. In this case, the rate of drug metabolism is directly proportional to the concentration of free drug and follows first-order kinetics. A constant percentage of the drug is metabolized over time, and the rate of elimination is proportional to the amount of drug in the body.

Most drugs used in the clinical setting are eliminated in this manner. A few drugs, such as aspirin, ethanol, and phenytoin, are used in higher doses, resulting in higher plasma concentrations. In these situations, C is much greater than Km, and the equation reduces to V(rate of drug metabolism) = (Vmax)(C)/(C) = Vmax. The enzyme system becomes saturated by a high free-drug concentration, and the rate of metabolism is constant over time. This is called zero-order kinetics, and a constant amount of drug is metabolized per unit of time. A large increase in serum concentration can result from a small increase in dose for drugs that follow zero-order elimination.

The liver is the principal organ of drug metabolism. Other organs that display considerable metabolic activity include the gastrointestinal tract, the lungs, the skin, and the kidneys. Following oral administration, many drugs are absorbed intact from the small intestine and transported to the liver via the portal system, where they are metabolized. This process is called first-pass metabolism, and may greatly limit the bioavailability of orally administered drugs. In general, all metabolic reactions can be classified as either Phase I or Phase II biotransformations. Phase I reactions usually convert the parent drug to a polar metabolite by introducing or unmasking a more polar site (-OH, $-NH_2$). If Phase I metabolites are sufficiently polar, they may be readily excreted. However, many Phase I metabolites undergo a subsequent reaction in which endogenous substances such as glucuronic acid, sulfuric acid, or an amino

acid combine with the metabolite to form a highly polar conjugate. Many drugs undergo these sequential reactions. However, Phase II reactions may precede Phase I reactions, as in the case of isoniazid.

Phase I reactions are usually catalyzed by enzymes of the cytochrome P450 system. These drug-metabolizing enzymes are located in the lipophilic membranes of the endoplasmic reticulum of the liver and other tissues. Three gene families, CYP1, CYP2, and CYP3, are responsible for most drug biotransformations. The CYP3A subfamily accounts for more than 50% of Phase I drug metabolism, predominantly by the CYP3A4 subtype. CYP3A4 is responsible for the metabolism of drugs commonly used in the intensive care setting, including acetaminophen, cyclosporine, diazepam, methadone, midazolam, spironolactone, and tacrolimus. Most other drug biotransformations are performed by CYP2D6 (e.g., clozapine, codeine, flecainide, haloperidol, oxycodone), CYP2C9 (e.g., phenytoin, S-warfarin), CYP2C19 (e.g., diazepam, omeprazole, propranolol), CYP2E1 (e.g., acetaminophen, enflurane, halothane), and CYP1A2 (e.g., acetaminophen, caffeine, theophylline, warfarin).

Drug biotransformation reactions may be enhanced or impaired by multiple factors, including age, enzyme induction or inhibition, pharmacogenetics, and the effects of other disease states [2]. For example, the metabolic pathways for acetaminophen have been well studied. Approximately 95% of the metabolism occurs via conjugation to glucuronide (50–60%) and sulfate (25–35%). Most of the remainder of acetaminophen is metabolized via the cytochrome P450 forming N-acetyl-p-benzoquinone imine (NAPQI), which is thought to be responsible for hepatotoxicity. This minor but important pathway is catalyzed by CYP 2E1, and to a lesser extent by CYP 1A2 and CYP 3A4. NAPQI is detoxified by reacting with either glutathione directly or through a glutathione transferase catalyzed reaction. When the hepatic synthesis of glutathione is overwhelmed, manifestations of toxicity appear, producing centrilobular necrosis. In the presence of a potent CYP 2E1 inhibitor, disulfiram, there was a 69% reduction in the urinary excretion of these 2E1 metabolic products, which supports the assignment of a major role for 2E1 in the formation of NAPQI [3]. Studies of inhibitors of other CYP pathways (e.g., 1A2 and 3A4) have failed to document a significant effect on the urinary excretion of glutathione conjugates [4]; thus, 2E1 appears to be the primary pathway overwhelmingly responsible for NAPQI. CYP 2E1 is unique among the CYP gene families in its ability to produce reactive oxygen radicals through a reduction of O_2, and is the only CYP system strongly induced by alcohol, which is itself a substrate. In addition to alcohol, isoniazid acts as an inducer and a substrate. Ketoconazole and other imidazole compounds are inducers but not substrates. Barbiturates and phenytoin, which are nonspecific inducers, have no role as CYP 2E1 inducers, nor are they substrates for that system. Phenytoin in fact may be hepatoprotective, because it is an inducer of the glucuronidation metabolic pathway for acetaminophen, thus shunting metabolism away from NAPQI production [5].

Elimination

Elimination is the process by which a drug is removed or "cleared" from the body. Clearance (CL) is usually referred to as the amount of blood from which all drug is removed per unit of time (volume/time). The main organs responsible for drug clearance are the kidneys and the liver. The total body clearance of a drug is equal to the sum of the clearances from all mechanisms. Typically, this is partitioned into renal and nonrenal clearance. Most elimination by the kidneys is accomplished by glomerular filtration. The amount of drug that is filtered is determined by glomerular integrity, the size and charge (electrostatic force of a molecule related

to whether it has gained or lost electrons, positive or negative respectively) of the drug, water solubility, and the extent of protein binding. Highly protein-bound drugs are not readily filtered. Therefore, estimation of the glomerular filtration rate (GFR) has traditionally served as an approximation of renal function.

In addition to glomerular filtration, drugs may be eliminated from the kidneys via active secretion. Secretion occurs predominantly at the proximal tubule, where active transport systems secrete primarily organic acids and bases. Organic acids include most cephalosporins, loop diuretics, methotrexate, nonsteroidal anti-inflammatories, penicillin, and thiazide diuretics. Organic bases include ranitidine and morphine. As drugs move toward the distal convoluting tubule, their concentration increases. High urine flow rates decrease the concentration of the drug in the distal tubule, decreasing the likelihood that the drug will diffuse from the lumen. For both weak acids and bases, the nonionized form of the drug is reabsorbed more readily. Altering the pH (ion trapping) can minimize reabsorption, by placing a charge on the drug and preventing its diffusion. For example, salicylate is a weak acid. In case of salicylate toxicity, urine alkalization places a charge on the molecule, and increases its elimination. The liver also contributes to elimination through metabolism or excretion into the bile. After a drug is secreted in the bile, it may then be either excreted into the feces or reabsorbed via enterohepatic recirculation [6].

The **half-life of elimination** is the time it takes to clear half of the drug from plasma. It is directly proportional to the Vd, and inversely proportional to CL: $t_{1/2\beta} = (0.693)\,(Vd)/CL$.

Special Populations

The term "special populations" as applied to drug development refers to discussions in the early 1990s among industry, academic, and regulatory scientists struggling with the then current practice that early drug development was focused predominantly on young, Caucasian male populations. Representatives from the US, Europe, and Japan jointly issued regulatory requirements for drug testing and labeling in "special populations" (namely the elderly) in 1993. In later discussions, this generalization was expanded to include four major demographic segments (women, the elderly, pediatric, and major ethnic groups); despite the large size of each of these population segments, pharmaceutical research had been limited in each of these areas. Current appreciation for these populations also benefits from a greater understanding of the heterogeneity of the eventual marketplace for many new chemical entities. More importantly, these "special populations" also represent diverse subpopulations of patients in whom dosing guidance is often needed, and likewise targeted clinical pharmacology research is essential.

Elderly

There are many physical signs consistent with aging, including wrinkles, change of hair color to gray or white, hair loss, lessened hearing, diminished eyesight, slower reaction times, and decreased agility. In clinical pharmacology, we are more concerned with how aging affects physiologic processes that dictate drug pharmacokinetics and pharmacodynamics. Advancing age is characterized by impairment in the function of the many regulatory processes that provide functional integration between cells and organs. Under these circumstances, failure to maintain homeostasis under conditions of physiologic stress can exist. This can often explain, at least in part, the increased interindividual variability that occurs as people age.

Cardiac structure and function, renal and gastrointestinal systems, and body composition are the physiologic systems most often implicated when pharmacokinetic or pharmacodynamic

differences are observed between elderly and young populations. Table 2.1 lists the primary physiologic factors affected by aging [7]. Recognition of these factors is important for predicting the implications of aging for drug pharmacokinetics especially.

With respect to absorption, the impact of age is unclear, and many conflicting results exist. While many studies have not shown significant age-related differences in absorption rates for specific drugs, the absorption of vitamin B_{12}, iron, and calcium is slower through reduced active transport mechanisms [8,9]. A reduction in first-pass metabolism is associated with aging, most likely due to a reduction in liver mass and blood flow. Likewise, drugs undergoing a significant first-pass effect experience an increase in bioavailability with age. This is the case for drugs like propranolol and labetalol. Conversely, drugs administered as prodrugs and requiring activation in the liver (e.g., ACE [angiotensin converting enzyme] inhibitors enalapril and perindopril) are likely to experience reduction in this phase and likewise reduced exposure of the active species.

Based on age-related changes in body composition, polar drugs, which are primarily water soluble, often exhibit smaller volumes of distribution, resulting in higher plasma concentrations in older patients. This is the case for agents including ethanol, theophylline, digoxin, and gentamicin [7,10]. Conversely, nonpolar compounds are often lipid soluble and exhibit larger volumes of distribution in older patients. The impact of the larger Vd is prolongation of half-life with age. This is the case for drugs such as chlormethiazole and thiopentone [11,12]. Conflicting results have been reported with

Table 2.1 Physiologic systems affected during aging that influence drug pharmacokinetic and/or pharmacodynamic behavior.

Physiologic System	Impact of Aging
Cardiac structure and function	• Reduced elasticity and compliance of the aorta and great arteries (higher systolic arterial pressure, increased impedance to left ventricular hypertrophy, and interstitial fibrosis) • Decrease in rate of myocardial relaxation • Left ventricle stiffens and takes longer to relax and fill in diastole • Isotonic contraction is prolonged and velocity of shortening reduced • Reduction in intrinsic heart rate and increased sinoatrial node conduction time
Renal system	• Renal mass decreases (reduction in nephrons) • Reduced blood flow in the afferent arterioles in the cortex • Renal plasma flow and glomerular filtration rate decline • Decrease in ability to concentrate the urine during water deprivation • Impaired response to water loading
Gastrointestinal system	• Secretion of hydrochloric acid and pepsin is decreased under basal conditions • Reduced absorption of several substances in the small intestine, including sugar, calcium, and iron • Decrease in lipase and trypsin secretion in the pancreas • Progressive reduction in liver volume and liver blood flow
Body composition	• Progressive reduction in total body water and lean body mass, resulting in a relative increase in body fat

respect to age effects on protein binding [13,14], making generalization difficult.

Several drug classes, including water-soluble antibiotics, diuretics, water-soluble beta-adrenoceptor blockers, and nonsteroidal anti-inflammatory drugs [7,15], exhibit changes in clearance with age as a result of declining renal function. With respect to hepatic metabolism, studies have shown that significant reductions in clearance with age are observed for Phase I pathways in the liver [16–18].

From the standpoint of a clinical trial, age categories are necessary to define the inclusion and exclusion criteria for the population targeted for enrollment. Most developed world countries have accepted the chronological age of 65 years as a definition of "elderly" or an older person. The salient point is that pharmaceutical sponsors are increasingly encouraged to include a broader range of ages in their pivotal trials than before, or specifically to target an elderly subpopulation in a separate trial, consistent with Food and Drug Administration (FDA) guidance. The FDA guideline for studies in the elderly is directed principally toward new molecular entities likely to have significant use in that population, either because the disease intended to be treated is characteristically a disease of aging (e.g., Alzheimer's disease) or because the population to be treated is known to include substantial numbers of geriatric patients (e.g., hypertension).

Pediatrics

As children develop and grow, changes in body composition, development of metabolizing enzymes, and maturation of renal and liver function all have impacts on drug disposition [19,20] (see also Chapter 22).

Renal

Renal function in the premature and full-term neonate, both glomerular filtration and tubular secretion, is significantly reduced compared to older children. Maturation of renal function is a dynamic process that begins during fetal life and is complete by early childhood. Maturation of tubular function is slower than that of glomerular filtration. The glomerular filtration rate is approximately 2–4 ml/minute/1.73 m^2 in full-term neonates, but it may be as low as 0.6–0.8 ml/minute/1.73 m^2 in preterm neonates. The glomerular filtration rate increases rapidly during the first two weeks of life and continues to rise until adult values are reached at 8–12 months of age. For drugs that are renally eliminated, impaired renal function decreases clearance, increasing the half-life. Therefore, for drugs that are primarily eliminated by the kidney, dosing should be performed in an age-appropriate fashion that takes into account both maturational changes in kidney function [21].

Hepatic

Hepatic biotransformation reactions are substantially reduced in the neonatal period. At birth, the cytochrome p450 system is 28% of that of the adult [22]. The expression of Phase I enzymes such as the P-450 cytochromes changes markedly during development. CYP3A7, the predominant CYP isoform expressed in the fetal liver, peaks shortly after birth and then declines rapidly to levels that are undetectable in most adults. Within hours after birth, CYP2E1 activity increases, and CYP2D6 becomes detectable soon thereafter. CYP3A4 and CYP2C appear during the first week of life, whereas CYP1A2 is the last hepatic CYP to appear, at 1–3 months of life [22,23]. The ontogeny of Phase II enzymes is less well established than the ontogeny of reactions involving Phase I enzymes. Available data indicate that the individual isoforms of glucuronosyltransferase (UGT) have unique maturational profiles with pharmacokinetic consequences. For example, the glucuronidation of acetaminophen (a substrate for UGT1A6 and, to a lesser extent, UGT1A9) is decreased in newborns and young children compared with adolescents and adults.

Glucuronidation of morphine (a UGT2B7 substrate) can be detected in premature infants as young as 24 weeks of gestational age [24,25].

Gastrointestinal

Overall, the rate at which most drugs are absorbed is slower in neonates and young infants than in older children. As a result, the time required to achieve maximal plasma levels is longer in the very young. The effect of age on enteral absorption is not uniform and is difficult to predict [19,23]. Gastric emptying and intestinal motility are the primary determinants of the rate at which drugs are presented to and dispersed along the mucosal surface of the small intestine. At birth, the coordination of antral contractions improves, resulting in a marked increase in gastric emptying during the first week of life. Similarly, intestinal motor activity matures throughout early infancy, with consequent increases in the frequency, amplitude, and duration of propagating contractions [26,27]. Changes in the intraluminal pH in different segments of the gastrointestinal tract can directly affect both the stability and the degree of ionization of a drug, thus influencing the relative amount of the drug available for absorption. During the neonatal period, intragastric pH is relatively elevated (greater than 4). Thus, oral administration of acid-labile compounds such as penicillin G produces greater bioavailability in neonates than in older infants and children [28]. In contrast, drugs that are weak acids, such as phenobarbital, may require larger oral doses in the very young in order to achieve therapeutic plasma levels. Other factors that affect the rate of absorption include age-associated development of villi, splanchnic blood flow, changes in intestinal microflora, and intestinal surface area [27].

Body Composition

Age-dependent changes in body composition alter the physiologic spaces into which a drug may be distributed. The percentage of total body water drops from about 85% in premature infants to 75% in full-term infants to 60% in the adult. Extracellular water decreases from 45% in the infant to 25% in the adult. Total body fat in the premature infant can be as low as 1%, compared to 15% in the normal, term infant. Many drugs are less bound to plasma proteins in the neonate and infant than in the older child [29]. Limited data in neonates suggest that the passive diffusion of drugs into the central nervous system is age dependent, as reflected by the progressive increase in the ratios of brain phenobarbital to plasma phenobarbital from 28 to 39 weeks of gestational age, demonstrating the increased transport of phenobarbital into the brain [30].

Pregnancy

The FDA classifies drugs into five categories of safety for use during pregnancy – that is, normal pregnancy, labor, and delivery – as outlined in Table 2.2 (see also Chapter 22). Few well-controlled studies of therapeutic drugs

Table 2.2 FDA categories of drug safety during pregnancy.

Category	Description
A	Controlled human studies show no fetal risks; these drugs are the safest
B	Animal studies show no risk to the fetus and no controlled human studies have been conducted, or animal studies show a risk to the fetus but well-controlled human studies do not
C	No adequate animal or human studies have been conducted, or adverse fetal effects have been shown in animals but no human data are available
D	Evidence of human fetal risk exists, but benefits may outweigh risks in certain situations (e.g., life-threatening disorders, serious disorders for which safer drugs cannot be used or are ineffective)
X	Proven fetal risks outweigh any possible benefit

have been conducted in pregnant women. Most information about drug safety during pregnancy is derived from animal studies and uncontrolled studies in people (e.g., postmarketing reports).

Observational studies have documented that pregnant women take a variety of medicines during pregnancy [31]. While changes in drug exposure during pregnancy are well documented, a mechanistic understanding of these effects is not clear [32]. The few studies that have been conducted suggest that bioavailability is not altered during pregnancy, though increased plasma volume and protein binding changes can alter the apparent volume of distribution of some drugs [32]. Likewise, changes in volume of distribution and clearance during pregnancy can cause increases or decreases in the terminal elimination half-life of drugs. Renal excretion of unchanged drugs is increased during pregnancy [32] and hence these agents may require dose increases. Likewise, the metabolism of drugs via select P450-mediated pathways (3A4, 2D6, and 2C9) and UGT isoenzymes is increased during pregnancy, necessitating increased dosages of drugs metabolized by these pathways [32,33]. In contrast, CYP1A2 and CYP2C19 activity is decreased during pregnancy, suggesting dosing reductions for agents metabolized via these pathways. The effect of pregnancy on transport proteins is unknown. These data are limited and, hence, more clinical evidence-based studies to determine the effect of pregnancy on the pharmacokinetics and pharmacodynamics of commonly used drugs are sorely needed (see also Chapter 22).

Organ Impairment

Renal Dysfunction

Renal failure can affect the pharmacokinetics of drugs. In renal failure, the binding of acidic drugs to albumin is decreased, because of competition with accumulated organic acids and uremia-induced structural changes in albumin, which decrease drug binding affinity, altering the volume of distribution [15]. Drugs that are more than 30% eliminated unchanged in the urine are likely to have significantly diminished clearance in the presence of renal insufficiency [15].

Hepatic Dysfunction

Drugs that undergo extensive first-pass metabolism may have a significantly higher oral bioavailability in patients with liver failure than in normal subjects. Gut hypomotility may delay the peak response to enterally administered drugs in these patients. Hypoalbuminemia or altered glycoprotein levels may affect the fractional protein binding of acidic or basic drugs, respectively. Altered plasma protein concentrations may affect the extent of tissue distribution of drugs that are normally highly protein bound. The presence of significant edema and ascites may alter the volume of distribution of highly water-soluble agents, such as aminoglycoside antibiotics. The capacity of the liver to metabolize drugs depends on hepatic blood flow and liver enzyme activity, both of which can be affected by liver disease. In addition, some p450 isoforms are more susceptible than others to liver disease, impairing drug metabolism [18].

Cardiac Dysfunction

Circulatory failure, or shock, can alter the pharmacokinetics of drugs frequently used in the intensive care setting. Drug absorption may be impaired because of bowel wall edema. Passive hepatic congestion may impede first-pass metabolism, resulting in higher plasma concentrations. Peripheral edema inhibits absorption by intramuscular parenteral routes. The balance of tissue hypoperfusion versus increased total body water with edema may unpredictably alter the volume of distribution. In addition, liver hypoperfusion may alter drug-metabolizing enzyme function, especially flow-dependent drugs such as lidocaine [34,35].

Drug Interactions

Patients are often treated with more than one drug, and often many, increasing the chance of a drug–drug interaction (see also Chapter 40). These interactions can alter absorption, distribution, metabolism, and clearance. Drug interactions can affect absorption through formation of drug–drug complexes (e.g., significantly increased bioavailability of fexofenadine in the presence of St. John's wort [36]), alterations in gastric pH, and changes in gastrointestinal motility. This can have a substantial impact on the bioavailability of enterally administered agents. The volume of distribution may be altered with competitive plasma protein binding and subsequent changes in free drug concentrations [13,14,37].

Drug biotransformation reactions vary greatly among individuals and are susceptible to drug–drug interactions. Induction is the process by which enzyme activity is increased by exposure to a certain drug, resulting in an increase in the metabolism of other drugs and lower plasma concentrations. Common inducers include barbiturates, carbamazepine, isoniazid, and rifampin. In contrast, inhibition is the process by which enzyme activity is decreased by exposure to a certain drug, resulting in a decrease in the metabolism of other drugs, and subsequent higher plasma concentrations. Common enzyme inhibitors include ciprofloxacin, fluconazole, metronidazole, quinidine, and valproic acid [2]. Inducers and inhibitors of Phase II enzymes have been less extensively characterized, but some clinical applications of this information have emerged, including the use of phenobarbital to induce glucuronyl transferase activity in icteric neonates. Water-soluble drugs are eliminated unchanged in the kidneys. The clearance of drugs that are excreted entirely by glomerular filtration is unlikely to be affected by other drugs. Organic acids and bases are renally secreted, and can compete with one another for elimination, resulting in unpredictable drug disposition [15].

Pharmacodynamics

Pharmacodynamics, in general terms, seeks to define what a drug does to the body (i.e., the effects or response to drug therapy). Pharmacodynamic modeling attempts to characterize measured, physiologic parameters before and after drug administration, with the effect defined as the change in a physiologic parameter relative to its pre-dose or baseline value. Baseline is defined as the physiologic parameter without drug dosing, and may be complicated in certain situations due to diurnal variations. Efficacy can be defined numerically as the expected sum of all beneficial effects following treatment (see also Chapter 33). In this case we refer to clinical and not necessarily economic benefits, though there clearly may be concordance (see also Chapter 34). Similarly, toxicity can be characterized either by the time course of a specific toxic event or the composite of toxic responses attributed to a common toxicity.

Overview

Pharmacodynamic response to drug therapy evolves only after active drug molecules reach their intended site(s) of action. Hence, the link between pharmacokinetic and pharmacodynamic processes is implicit. Likewise, the respective factors that influence various subprocesses (absorption, distribution, tolerance, etc.) are relevant and may necessitate separate study. Differences in pharmacodynamic time course among drug entities can be broadly associated with the nature of the concentration–effect relationship as being direct (effect is directly proportional to concentration at the site of measurement, usually the plasma) or indirect (effect exhibits some type of temporal delay with respect to drug concentration, either because of differences between site of action and measurement, or because the effect

of interest results after other physiologic or pharmacologic conditions are satisfied).

Direct effect relationships are easily observed with cardiovascular agents whose site of action is the vascular space. Pharmacologic effects such as blood pressure, ACE inhibition, and inhibition of platelet aggregation can be characterized by direct response relationships. Such relationships can usually be defined by three typical patterns: linear, hyperbolic (E_{max}), and sigmoid E_{max} functions [38]. These are shown in Figure 2.3. In each case, the plasma concentration and drug concentration at the effect site are proportional. Likewise, the concentration–effect relationship is assumed to be independent of time.

Other drugs exhibit an indirect relationship between concentration and response. In this case, the concentration–effect relationship is time dependent. One explanation for such effects is hysteresis, which refers to the phenomenon where there is a time lapse between a cause and its effect. With respect to pharmacodynamics, this most often indicates a situation in which there is a delay in equilibrium between plasma drug concentration and the concentration of active substance at the effect site (thiopental,

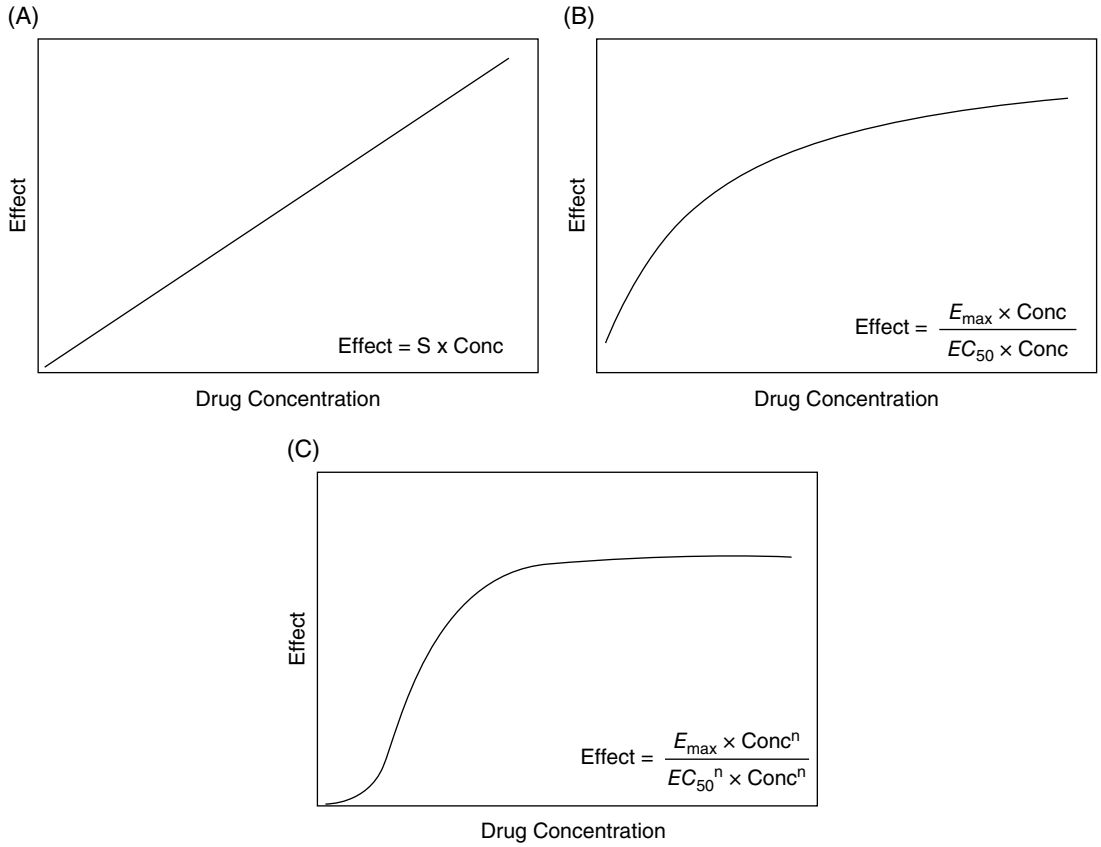


Figure 2.3 Representative pharmacodynamic relationships for drugs that exhibit direct responses: (A) linear, (B) hyperbolic, and (C) Sigmoid-E_{max} relationships are shown. S is the slope of the linear response; E_{max} refers to the maximum effect observed; EC_{50} refers to the concentration at which 50% of the maximal response is achieved; and n is the degree of sigmoidicity or shape factor (sometimes referred to as the Hill coefficient).

fentanyl, and many others). Three conditions predominate: the biophase (actual site of drug action) is not in the central compartment, the mechanism of action involves protein synthesis, and/or active metabolites are present. One can conceptualize a hypothetical effect compartment (a physical space where drug concentrations are directly correlated with drug actions) such that the relationships defined in Figure 2.4 are only observed when the effect site concentration (Ce) is used as opposed to the plasma concentration (Cp). In this situation, a hysteresis loop is observed when plotting Ce versus Cp (see Figure 2.4).

More complicated models (indirect response models) have been used to express the same observations, but typically necessitate a greater understanding of the underlying physiologic process (e.g., cell trafficking, enzyme recruitment, etc.) [38]. The salient point is that pharmacodynamic characterization and likewise dosing guidance derived from such investigation stand to be more informative than drug concentrations alone. Likewise, pharmacodynamics may be the discriminating characteristic that defines dose adjustment in special populations. This is the case for the observed markedly enhanced sensitivity in infants compared with

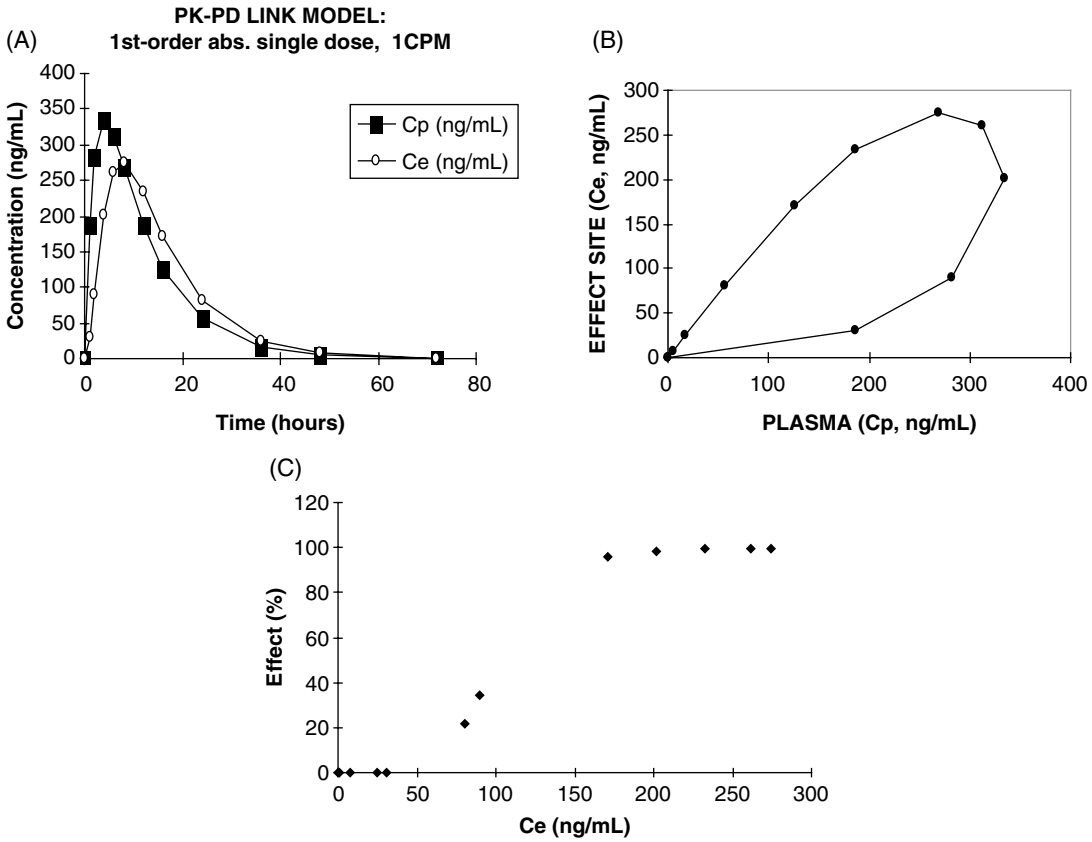


Figure 2.4 (A) Concentration–time, (B) hysteresis, and (C) effect–concentration plots, illustrating the use of an effect compartment to explain observed hysteresis. Ce, concentration at the effect site; Cp, plasma concentration; 1CPM, one-compartment model; PD, pharmacodynamics; PK, pharmacokinetics.

older children and adults with respect to the immunosuppressive effects of cyclosporine [39], and for calcium channel blocking effects on the PR interval in the elderly [40,41].

Pharmacogenomics

Pharmacogenomics is the study of how an individual's genetic inheritance affects the body's response to drugs (see also Chapter 30). Pharmacogenomics holds the promise that drugs might one day be tailored to individuals and adapted to each person's own genetic makeup. Environment, diet, age, lifestyle, and state of health all can influence a person's response to medicines, but understanding an individual's genetic composition is thought to be the key to creating personalized drugs with greater efficacy and safety. Pharmacogenomics combines traditional pharmaceutical sciences, such as biochemistry, with annotated knowledge of genes, proteins, and single nucleotide polymorphisms (SNPs). Genetic variations, or SNPs, in the human genome can be a diagnostic tool to predict a person's drug response. For SNPs to be used in this way, a person's DNA must be sequenced for the presence of specific SNPs. SNP screenings will benefit drug development and those people whose pharmacogenomic screening shows that the drug being tested would be harmful or ineffective for them would be excluded from clinical trials. Prescreening clinical trial subjects might also allow clinical trials to be smaller, faster, and therefore less expensive. Finally, the ability to assess an individual's reaction to a drug before it is prescribed will increase confidence in prescribing the drug and the patient's confidence in taking the drug, which in turn should encourage the development of new drugs tested in a like manner. For example, the major enzyme responsible for tacrolimus metabolism is CYP3A. CYP3A5 genes have multiple SNPs. One study found that at 3, 6, and 12 months after heart transplantation, there was a significant difference in tacrolimus blood concentrations per dose/kg/day between the CYP3A5 *1/*3 (CYP3A5 expresser) and the *3/*3 (nonexpresser) genotypes, with the *1/*3 patients requiring larger tacrolimus doses to achieve the same blood concentration. It was concluded that specific genotypes of CYP3A5 in pediatric heart transplant patients require larger tacrolimus doses to maintain their tacrolimus blood concentration, and that this information could be used prospectively to manage patients' immunosuppressive therapy. (See also Chapter 30 for molecular pharmacoepidemiology.)

Model-Informed Drug Development

One of the more recent developments in the evolution of clinical pharmacology in the facilitation of early stage drug development is in the implementation of model-informed drug development (MIDD) principles by many pharmaceutical and biotech companies. The approach is also endorsed by the global regulatory community, including the European Medicines Agency (EMA) and the FDA. The use of modeling and simulation approaches to derisk decision making in drug development is not new, but the systematic integration of the unique model assets in an evolving computing environment that expands with knowledge about candidate molecules and/or vaccines is still a work in progress for many pharmaceutical sponsors. However, feedback from early adopters suggests that the approach can reduce both time and cost in drug development when conducted in an appropriate manner. Figure 2.5 highlights many of the common early drug development decision milestones, in conjunction with the various model types and methodologies that represent key stage-gate milestones.

Many of these milestones represent contributions from clinical pharmacology and the

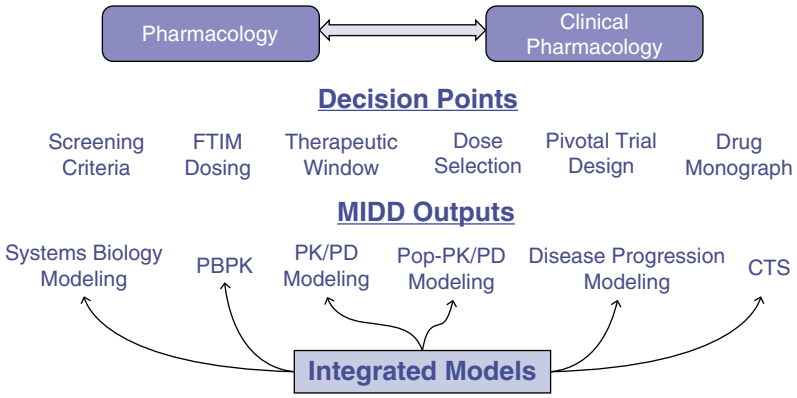


Figure 2.5 The model-informed drug development (MIDD) approach learns and confirms key characteristics of new molecular entities in a quantitative manner, with the goal of providing explicit, reproducible, and predictive evidence for optimizing drug development plans and enabling critical decisions. CTS, clinical trial simulation; FTIM, first time in man; PBPK, physiologically based pharmacokinetics; PD, pharmacodynamics; PK, pharmacokinetics.

supportive quantitative disciplines which collaborate in the MIDD effort (e.g., bioinformatics, system pharmacology, drug metabolism and pharmacokinetics [DMPK], pharmacometrics, and biostatistics). As Figure 2.5 suggests, many of these milestones are not only critical to the progression of drug and/or vaccine candidates, but also represent critical go/no go criteria requiring quantitative definition around the pace of potential outcomes. The MIDD likewise is effective at generating scenarios that explore the space of potential outcomes, either through direct experimentation or model-based projection (i.e., simulation).

In addition to the utility of MIDD in the decision-making process, MIDD implementation generates modeling assets that can be used in later stages of drug development. These can represent inputs to epidemiologic modeling and simulation exercises that explore the utility of projecting candidate attributes on target populations of interest, and also accommodate the complexity of the existing standard of care, population, and subpopulation differences influenced by socioeconomic and lifestyle factors. This represents a new frontier for these disciplines to further interact and inform each other.

Conclusion

Clinical pharmacology serves an important role in the development of new drugs and the management of pharmacotherapy. It is essential knowledge that must inform the drug developer, the investigator or trialist, the regulator, and the caregiver in their respective settings. In the context of pharmacoepidemiologic investigations, clinical pharmacology also provides a fundamental backbone for understanding the expected associations between drug therapy and clinical benefit, as well as potential toxicity. The pharmacoepidemiologist must also have intimate knowledge of clinical pharmacology, as the projection of performance (clinical and economic), the connection between utilization, compliance, and the complexities of multimodal therapy, and the associations of drug behavior with disease- or population-specific indices must be defined relative to the known clinical pharmacologic principles that govern how we expect drugs to behave in humans. In an era in which more holistic approaches are sought to maintain homeostasis and clinical strategies engage more preventive approaches, clinical pharmacology will be an essential discipline to discriminate options that are truly beneficial.

References

1 Wagner JG. *Biopharmaceutics and Relevant Pharmacokinetics*. Hamilton, Canada: Drug Intelligence Publications, 1971.

2 Woolf TF, ed. Handbook of Drug Metabolism. New York: Marcel Dekker, 1999.

3 Chen W, Koenigs LL, Thompson SJ, *et al*. Oxidation of acetaminophen to its toxic quinone imine and nontoxic catechol metabolites by baculovirus-expressed and purified human cytochromes P450 2E1 and 2A6. *Chem Res Toxicol* 1998; **11**(4): 295–301.

4 Manyike PT, Kharasch ED, Kalhorn TF, Slattery JT. Contribution of CYP2E1 and CYP3A to acetaminophen reactive metabolite formation. *Clin Pharmacol Ther* 2000; **67**(3): 275–82.

5 Kostrubsky SE, Sinclair JF, Strom SC, *et al*. Phenobarbital and phenytoin increased acetaminophen hepatotoxicity due to inhibition of UDP-glucuronosyltransferases in cultured human hepatocytes. *Toxicol Sci* 2005; **87**(1): 146–55.

6 Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd edn. New York: Marcel Dekker, 1982.

7 Mangoni AA. Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. *Br J Clin Pharmacol* 2004; **57**(1): 6–14.

8 Montgomery RD, Haeney MR, Ross IN, *et al*. The ageing gut: a study of intestinal absorption in relation to nutrition in the elderly. *Q J Med* 1978; **47**(186): 197–24.

9 Saltzman JR, Russell RM. The aging gut: nutritional issues. *Gastroenterol Clin North Am* 1998; **27**(2): 309–24.

10 Bhutto A, Morley JE. The clinical significance of gastrointestinal changes with aging. *Curr Opin Clin Nutr Metab Care* 2008; **11**(5): 651–60.

11 Stanski DR, Maitre PO. Population pharmacokinetics and pharmacodynamics of thiopental: the effect of age revisited. *Anesthesiology* 1990; **72**(3): 412–22.

12 Fagan D, Lamont M, Jostell KG, Tiplady B, Scott DB. A study of the psychometric effects of chlormethiazole in healthy young and elderly subjects. *Age Ageing* 1990; **19**(6): 395–402.

13 Karim A, Noveck R, McMahon FG, *et al*. Oxaprozin and piroxicam, nonsteroidal antiinflammatory drugs with long half-lives: effect of protein-binding differences on steady-state pharmacokinetics. *J Clin Pharmacol* 1997; **37**(4): 267–78.

14 Wallace SM, Verbeeck RK. Plasma protein binding of drugs in the elderly. *Clin Pharmacokinet* 1987; **12**(1): 41–72.

15 Tett SE, Kirkpatrick CM, Gross AS, McLachlan AJ. Principles and clinical application of assessing alterations in renal elimination pathways. *Clin Pharmacokinet* 2003; **42**(14): 1193–211.

16 Cornet M, Mertens K, Callaerts A, Sonck W, Vercruysse A, Rogiers V. Age- and gender-related changes in the hepatic metabolism of 2-methylpropene and relationship to epoxide metabolizing enzymes. *Mech Ageing Dev* 1994; **74**(1–2): 103–15.

17 Henderson GI, Speeg KV Jr., Roberts RK, Perez A, Schenker S. Effect of aging on hepatic elimination of cimetidine and subsequent interaction of aging and cimetidine on aminopyrine metabolism. *Biochem Pharmacol* 1988; **37**(13): 2667–73.

18 Durnas C, Loi CM, Cusack BJ. Hepatic drug metabolism and aging. *Clin Pharmacokinet* 1990; **19**(5): 359–89.

19 Strolin Benedetti M, Baltes EL. Drug metabolism and disposition in children. *Fundam Clin Pharmacol* 2003; **17**(3): 281–99.

20 Meibohm B, Laer S, Panetta JC, Barrett JS. Population pharmacokinetic studies in pediatrics: issues in design and analysis. *Aaps J* 2005; **7**(2): E475–87.

21 Johnson TN. The problems in scaling adult drug doses to children. *Arch Dis Child* 2008; **93**(3): 207–11.

22 Blake MJ, Castro L, Leeder JS, Kearns GL. Ontogeny of drug metabolizing enzymes in the neonate. *Semin Fetal Neonatal Med* 2005; **10**(2): 123–38.

23 Johnson TN, Thomson M. Intestinal metabolism and transport of drugs in children: the effects of age and disease. *J Pediatr Gastroenterol Nutr* 2008; **47**(1): 3–10.

24 Allegaert K, Vanhaesebrouck S, Verbesselt R, van den Anker JN. In vivo glucuronidation activity of drugs in neonates: extensive interindividual variability despite their young age. *Ther Drug Monit* 2009; **31**(4): 411–15.

25 Knibbe CA, Krekels EH, van den Anker JN, *et al*. Morphine glucuronidation in preterm neonates, infants and children younger than 3 years. *Clin Pharmacokinet* 2009; **48**(6): 371–85.

26 Riezzo G, Indrio F, Raimondi F, *et al*. Maturation of gastric electrical activity, gastric emptying and intestinal permeability in preterm newborns during the first month of life. *Ital J Pediatr* 2009; **35**(1): 6.

27 Lebenthal A, Lebenthal E. The ontogeny of the small intestinal epithelium. *JPEN J Parenter Enteral Nutr* 1999; **23**(5 Suppl): S3–6.

28 Bergdahl S, Eriksson M, Finkel Y, Lannergren K. Oral absorption of flucloxacillin in infants and young children. *Acta Pharmacol Toxicol (Copenh)* 1986; **58**(4): 255–8.

29 Aranda JV, Varvarigou A, Beharry K, *et al*. Pharmacokinetics and protein binding of intravenous ibuprofen in the premature newborn infant. *Acta Paediatr* 1997; **86**(3): 289–93.

30 Onishi S, Ohki Y, Nishimura Y, *et al*. Distribution of phenobarbital in serum, brain and other organs from pediatric patients. *Dev Pharmacol Ther* 1984; **7**(3): 153–9.

31 Dawes M, Chowienczyk PJ. Drugs in pregnancy: pharmacokinetics in pregnancy. *Best Pract Res Clin Obstet Gynaecol* 2001; **15**(6): 819–26.

32 Anderson GD. Pregnancy-induced changes in pharmacokinetics: a mechanistic-based approach. *Clin Pharmacokinet* 2005; **44**(10): 989–1008.

33 Chen H, Yang K, Choi S, Fischer JH, Jeong H. Up-regulation of UDP-glucuronosyltransferase (UGT) 1A4 by 17beta-estradiol: a potential mechanism of increased lamotrigine elimination in pregnancy. *Drug Metab Dispos* 2009; **37**(9): 1841–7.

34 Shammas FV, Dickstein K. Clinical pharmacokinetics in heart failure: an updated review. *Clin Pharmacokinet* 1988; **15**(2): 94–113.

35 Sato, T, Ono T, Ozaki M, Sakio H, Yasuda S. [Lidocaine metabolism associated with epidural anesthesia in patients for hepatic surgery]. *Masui* 1999; **48**(12): 1318–22.

36 Wang Z, Hamman MA, Huang SM, Lesko LJ, Hall SD. Effect of St John's wort on the pharmacokinetics of fexofenadine. *Clin Pharmacol Ther* 2002; **71**(6): 414–20.

37 Bruguerolle B. Chronopharmacokinetics: current status. *Clin Pharmacokinet* 1998; **35**(2): 83–94.

38 Rohatagi S, Martin N, Barrett JS. Pharmacokinetic/pharmacodynamic modeling in drug development. In: Krishna R., ed. *Applications of Pharmacokinetic Principles in Drug Development*. Hingham, MA: Kluwer Academic/Plenum, 2004, pp. 333–72.

39 Marshall JD, Kearns GL. Developmental pharmacodynamics of cyclosporine. *Clin Pharmacol Ther* 1999; **66**(1): 66–75.

40 Johnston DL, Lesoway R, Humen DP, Kostuk WJ. Clinical and hemodynamic evaluation of propranolol in combination with verapamil, nifedipine and diltiazem in exertional angina pectoris: a placebo-controlled, double-blind, randomized, crossover study. *Am J Cardiol* 1985; **55**(6): 680–7.

41 De Cicco M, Macor F, Robieux I, *et al*. Pharmacokinetic and pharmacodynamic effects of high-dose continuous intravenous verapamil infusion: clinical experience in the intensive care unit. *Crit Care Med* 1999; **27**(2): 332–9.

3

Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

Pharmacoepidemiology applies the methods of epidemiology to the content area of clinical pharmacology. Chapter 2 reviewed the basic principles of clinical pharmacology. Therefore, in order to understand the approaches and methodologic issues specific to the field of pharmacoepidemiology, the basic principles of epidemiology must be understood as well. To that end, this chapter will begin with an overview of the scientific method in general. This will be followed by a discussion of the different types of errors one can make in designing a study. Next, the chapter will review the criteria for the causal nature of an association, which are how one can decide whether an association demonstrated in a particular study is, in fact, a causal association. Finally, the specific study designs available for epidemiologic studies, or in fact for any clinical studies, will be reviewed. The next chapter discusses a specific methodologic issue which needs to be addressed in any study, but which is of particular importance for pharmacoepidemiologic studies: the issue of sample size. These two chapters are intended to be an introduction to the field of epidemiology for the neophyte. More information on these principles can be obtained from any textbook on epidemiology or clinical epidemiology [1–24].

Overview of the Scientific Method

The scientific method to investigate a research question involves a three-stage process (see Figure 3.1). In the first stage, one selects a group of subjects for study. These subjects may be patients or animals or biologic cells, and are the sources for the data sought by the study to answer a question of interest. Second, one uses the information obtained in this sample of study subjects to generalize and draw a conclusion about a population in general. This conclusion is referred to as an association. Third, one generalizes again, drawing a conclusion about a scientific theory or causation. Each will be discussed in turn.

Any given study is performed on a selection of individuals, who represent the *study subjects*. These study subjects should theoretically represent a random sample of some defined population. For example, one might perform a randomized clinical trial of the efficacy of enalapril in lowering blood pressure, randomly allocating a total of 40 middle-aged hypertensive men to receive either enalapril or placebo and observing their blood pressure six weeks later. One might expect to see the blood

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

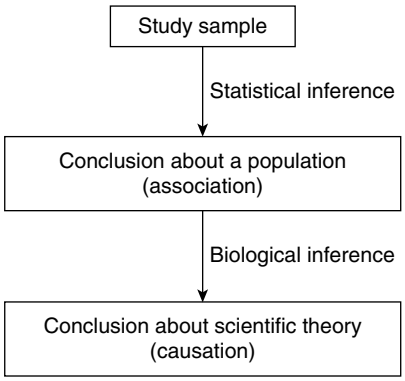


Figure 3.1 Overview of the scientific method.

pressure of the 20 men treated with the active drug decrease more than the blood pressure of the 20 men treated with a placebo. In this example, the 40 study subjects would represent the study sample, theoretically a random sample of middle-aged hypertensive men. In reality, the study sample is almost never a true random sample of the underlying target population, because it is logistically impossible to identify every individual who belongs in the target population and then randomly choose from among them. However, the study sample is usually treated as if it were a random sample of the target population.

At this point, one would be tempted to make a generalization that enalapril lowers blood pressure in middle-aged hypertensive men. However, one must explore whether this observation could have occurred simply by chance; that is, due to random variation. If the observed outcome in the study was simply a chance occurrence, then the same observation might not have been seen if one had chosen a different sample of 40 study subjects. Perhaps more importantly, it might not exist if one were able to study the entire theoretical population of all middle-aged hypertensive men. In order to evaluate this possibility, one can perform a statistical test, which allows an investigator to quantitate the probability that the observed outcome in this study (i.e., the difference seen between the two study groups) could have happened simply by chance. There are explicit rules and procedures for how one should properly make this determination: the science of statistics. If the results of any study under consideration demonstrate a "statistically significant difference" (i.e., ruling out the probability of a chance occurrence), then one is said to have an *association*. The process of assessing whether random variation could have led to a study's findings is referred to as *statistical inference*, and represents the major role for statistical testing in the scientific method.

If there is no statistically significant difference, then the process in Figure 3.1 stops. If there is an association, then one is tempted to generalize the results of the study even further, to state that enalapril is an antihypertensive drug in general. This is referred to as *scientific or biologic inference*, and the result is a conclusion about *causation*, that the drug really does lower blood pressure in a population of treated patients. To draw this type of conclusion, however, requires one to generalize to populations other than that included in the study, including types of people who were not represented in the study sample, such as women, children, and the elderly. Although it may be apparent in this example that this is in fact appropriate, that may well not always be the case. Unlike statistical inference, there are no precise quantitative rules for biologic inference. Rather, one needs to examine the data at hand in light of all other relevant data in the rest of the scientific literature, and make a subjective judgment. To assist in making that judgment, however, one can use the criteria for the causal nature of an association described later in the chapter. First, however, we will place causal associations into proper perspective by describing the different types of errors that can be made in performing a study and the different types of associations in which each results.

Types of Errors That One Can Make in Performing a Study

There are four basic types of associations that can be observed in a study (Table 3.1). The basic purpose of research is to differentiate among them.

First, of course, one could have no association.

Second, one could have an *artifactual association*; that is, a spurious or false association. This can occur by either of two mechanisms: chance or bias. Chance is unsystematic, or random, variation. The purpose of statistical testing in science is to evaluate this, estimating the probability that the result observed in a study could have happened purely by chance.

The other possible mechanism for creating an artifactual association is bias. Epidemiologists' use of the term bias is different from that of the lay public. To an epidemiologist, *bias* is systematic variation, a consistent manner in which two study groups are treated or evaluated differently. This consistent difference can create an apparent association where one actually does not exist. Of course, it also can mask a true association.

There are many different types of potential biases [25]. For example, consider an interview study in which the research assistant is aware of the investigator's hypothesis. Attempting to please the boss, the research assistant might probe more carefully during interviews with one study group than during interviews with the other. This difference in how carefully the interviewer probes could create an apparent but false association, which is referred to as interviewer bias. Another example would be a study of drug-induced birth defects that compares children with birth defects to children without birth defects. A mother of a child with birth defect, when interviewed about any drugs she took during her pregnancy, may be likely to remember drug ingestion during pregnancy with greater accuracy than a mother of a healthy child, because of the unfortunate experience she has undergone. The improved recall in the mothers of the children with birth defects may result in false apparent associations between drug exposure and birth defects. This systematic difference in recall is referred to as recall bias [26].

Table 3.1 Types of associations between factors under study.

1) None (independent)
2) Artifactual (spurious or false)
a) Chance (unsystematic variation)
b) Bias (systematic variation)
3) Indirect (confounded)
4) Causal (direct or true)

Note that biases, once present, cannot be corrected. They represent errors in the study design that can result in incorrect results in the study. It is important to note that a *statistically significant result is no protection against a bias*; one can have a very precise measurement of an incorrect answer! The only protection against biases is proper study design. (See Chapter 43 for more discussion about biases in pharmacoepidemiologic studies.)

Third, one can have an indirect, or confounded, association. A *confounding variable*, or *confounder*, is a variable, other than the risk factor and other than the outcome under study, which is related independently to both the risk factor and the outcome and which may create an apparent association or mask a real one. For example, a study of risk factors for lung cancer could find a very strong association between having yellow fingertips and developing lung cancer. This is obviously not a causal association, but an indirect association, confounded by cigarette smoking. Specifically, cigarette smoking causes both yellow fingertips and lung cancer. Although this example is transparent, most examples of confounding are not. In designing a study, one must consider every variable that can be associated with the risk factor under study or the outcome variable under study, in order to

Table 3.2 Approaches to controlling confounding.

1) Random allocation
2) Subject selection
a) Exclusion
b) Matching
3) Data analysis
a) Stratification
b) Mathematical modeling

plan to deal with it as a potential confounding variable. Preferably, one will be able to specifically control for the variable, using one of the techniques listed in Table 3.2. (See Chapters 33 and 43 for more discussion about confounding in pharmacoepidemiologic studies.)

Fourth, and finally, there are true, causal associations.

Thus, there are three possible types of errors that can be produced in a study: random error, bias, and confounding. The probability of random error can be quantitated using statistics. Bias needs to be prevented by designing the study properly. Confounding can be controlled either in the design of the study or in its analysis. If all three types of errors can be excluded, then one is left with a true, causal association.

Criteria for the Causal Nature of an Association

The "criteria for the causal nature of an association" were first put forth by Sir Austin Bradford Hill [27], but have been described in various forms since, each with some modification. Probably the best known description of them was in the first Surgeon General's Report on Smoking and Health [28], published in 1964. These criteria are presented in Table 3.3, in no particular order. No one of them is absolutely necessary for an association to be a causal association. Analogously, no one of them is sufficient for an association to be considered a causal association. Essentially, the more criteria that are present, the more likely it is that an association is a causal association. The fewer criteria that are met, the less likely it is that an association is a causal association. Each will be discussed in turn.

Table 3.3 Criteria for the causal nature of an association.

1) Coherence with existing information (biologic plausibility)
2) Consistency of the association
3) Time sequence
4) Specificity of the association
5) Strength of the association
a) Quantitative strength
b) Dose–response relationship
c) Study design

The first criterion listed in Table 3.3 is *coherence with existing information* or *biological plausibility*. This refers to whether the association makes sense, in light of other types of information available in the literature. These other types of information could include data from other human studies, data from studies of other related questions, data from animal studies, or data from *in vitro* studies, as well as scientific or pathophysiologic theory. To use the example provided earlier, it clearly was not biologically plausible that yellow fingertips could cause lung cancer, and this provided the clue that confounding was present. Using the example of the association between cigarettes and lung cancer, cigarette smoke is a known carcinogen, based on animal data. In humans, it is known to cause cancers of the head and neck, the pancreas, and the bladder. Cigarette smoke also goes down into the lungs, directly exposing the tissues in question. Thus, it certainly is biologically plausible that cigarettes could *cause* lung cancer [29]. It is much more reassuring if an association found in a particular study makes sense, based on previously available information, and this leads one to be more comfortable that it

might be a causal association. Clearly, however, one could not require that this criterion always be met, or one would never have a major breakthrough in science.

The second criterion listed in Table 3.3 is the *consistency of the association*. A hallmark of science is reproducibility: if a finding is real, one should be able to reproduce it in a different setting. This could include different geographic settings, different study designs, different populations, and so on. For example, in the case of cigarettes and lung cancer, the association has now been reproduced in many different studies, in different geographic locations, using different study designs [30]. The need for reproducibility is such that one should never believe a finding reported only once; there may have been an error committed in the study, which is not apparent to either the investigator or the reader.

The third criterion listed is the *time sequence*: a cause must precede an effect. Although this may seem obvious, there are study designs from which this cannot be determined. For example, if one were to perform a survey in a classroom of 200 medical students, asking each if they were currently taking diazepam and also whether they were anxious, one would find a strong association between the use of diazepam and anxiety, but this does not mean that diazepam causes anxiety! Although this is obvious, as it is not a biologically plausible interpretation, one cannot differentiate from this type of cross-sectional study which variable came first and which came second. In the example of cigarettes and lung cancer, obviously the cigarette smoking usually precedes the lung cancer, as a patient would not survive long enough to smoke much if the opposite were the case.

The fourth criterion listed in Table 3.3 is *specificity*. This refers to the question of whether the cause ever occurs without the presumed effect, and whether the effect ever occurs without the presumed cause. This criterion is almost never met in biology, with the occasional exception of infectious diseases. Measles never occurs without the measles virus, but even in this example, not everyone who becomes infected with the measles virus develops clinical measles. Certainly, not everyone who smokes develops lung cancer, and not everyone who develops lung cancer was a smoker. This is one of the major points the tobacco industry stresses when it attempts to make the claim that cigarette smoking has not been proven to cause lung cancer. Some authors even omit this as a criterion, as it is so rarely met. When it is met, however, it provides extremely strong support for a conclusion that an association is causal.

The fifth criterion listed in Table 3.3 is the *strength of the association*. This includes three concepts: its quantitative strength, dose–response, and the study design. Each will be discussed in turn.

The *quantitative strength* of an association refers to the effect size. To evaluate this, one asks whether the magnitude of the observed difference between the two study groups is large. A quantitatively large association can only be created by a causal association or a large error, which should be apparent in evaluating the methods of a study. A quantitatively small association may still be causal, but it could be created by a subtle error, which would not be apparent in evaluating the study. Conventionally, epidemiologists consider an association with a relative risk of less than 2.0 a weak association. Certainly, the association between cigarette smoking and lung cancer is a strong association: studies show relative risks ranging between 10.0 and 30.0 [30].

A *dose–response relationship* is an extremely important and commonly used concept in clinical pharmacology and is used similarly in epidemiology. It exists when an increase in the intensity of an exposure results in an increased risk of the disease under study. Equivalent to this is a *duration–response relationship*, which exists when a longer exposure causes an increased risk of the disease. The presence of either relationship strongly implies that an

association is, in fact, a causal association. Certainly in the example of cigarette smoking and lung cancer, it has been shown repeatedly that an increase in either the number of cigarettes smoked each day or in the number of years of smoking increases the risk of developing lung cancer [30].

Finally, *study design* refers to two concepts: whether the study was well designed, and which study design was used in the studies in question. The former refers to whether the study was subject to one of the three errors described earlier in this chapter, namely random error, bias, and confounding. Table 3.4 presents the study designs typically used for epidemiologic studies, or in fact for any clinical studies. They are organized in a hierarchical fashion. As one advances from the designs at the bottom of the

Table 3.4 Advantages and disadvantages of epidemiologic study designs.

Study design	Advantages	Disadvantages
Randomized clinical trial (experimental study)	Most convincing design Only design which controls for unknown or unmeasurable confounders	Most expensive Artificial Logistically most difficult Ethical objections
Cohort study	Can study multiple outcomes Can study uncommon exposures Selection bias less likely Unbiased exposure data Incidence data available	Possibly biased outcome data More expensive If done prospectively, may take years to complete
Case–control study	Can study multiple exposures Can study uncommon diseases Logistically easier and faster Less expensive	Control selection problematic Possibly biased exposure data
Analyses of secular trends	Can provide rapid answers	No control of confounding
Case series	Easy quantitation of incidence	No control group, so cannot be used for hypothesis testing
Case reports	Cheap and easy method for generating hypotheses	Cannot be used for hypothesis testing

table to those at the top, studies get progressively harder to perform, but are progressively more convincing. In other words, associations shown by studies using designs at the top of the list are more likely to be causal associations than associations shown by studies using designs at the bottom of the list. The association between cigarette smoking and lung cancer has been reproduced in multiple well-designed studies, using analyses of secular trends, case–control studies, and cohort studies. However, it has not been shown using a randomized clinical trial, which is the "Cadillac" of study designs, as will be discussed later in the chapter. This is the other major defense the tobacco industry employs. Of course, it would not be ethical or logistically feasible to randomly allocate individuals to smoke or not to smoke and expect to follow them for 20 years to observe the outcome in each group.

The issue of causation is discussed more in Chapter 10 as it relates to the process of spontaneous reporting of adverse drug reactions, and in Chapter 29 as it relates to determining causation in case reports.

Epidemiologic Study Designs

In order to clarify the concept of study design further, each of the designs in Table 3.4 will be discussed in turn, starting at the bottom of the list and working upward.

Case Reports

Case reports are simply reports of events observed in single patients. As used in pharmacoepidemiology, a case report describes a single patient who was exposed to a drug and experiences a particular, usually adverse, outcome. For example, one might see a published case report about a young woman who was taking oral contraceptives and who suffered a pulmonary embolism.

Case reports are useful for raising hypotheses about drug effects, to be tested with more rigorous study designs. However, in a case report one cannot know if the patient reported is either typical of those with the exposure or typical of those with the disease. Certainly, one cannot usually determine whether the adverse outcome was due to the drug exposure or would have happened anyway. As such, it is very rare that a case report can be used to make a statement about causation. One exception to this would be when the outcome is so rare and so characteristic of the exposure that one knows that it was likely to be due to the exposure, even if the history of exposure were unclear. An example of this is clear cell vaginal adenocarcinoma occurring in young women exposed *in utero* to diethylstilbestrol [31]. Another exception would be when the disease course is very predictable and the treatment causes a clearly apparent change in this disease course. An example would be the ability of penicillin to cure streptococcal endocarditis, a disease that is nearly uniformly fatal in the absence of treatment. Case reports can be particularly useful to document causation when the treatment causes a change in disease course which is reversible, such that the patient returns to their untreated state when the exposure is withdrawn, can be treated again, and when the change returns upon repeat treatment. Consider a patient who is suffering from an overdose of methadone, a long-acting narcotic, and is comatose. If this patient is then treated with naloxone, a narcotic antagonist, and immediately awakens, this would be very suggestive that the drug indeed is efficacious as a narcotic antagonist. As the naloxone wears off the patient would become comatose again, and then if they were given another dose of naloxone they would awaken again. This, especially if repeated a few times, would represent strong evidence that the drug is indeed effective as a narcotic antagonist. This type of challenge–rechallenge situation is relatively uncommon, however, as physicians generally will avoid

exposing a patient to a drug if the patient experienced an adverse reaction to it in the past. This issue is discussed in more detail in Chapters 10 and 29.

Case Series

Case series are collections of patients, all of whom have a single exposure, whose clinical outcomes are then evaluated and described. Often they are from a single hospital or medical practice. Alternatively, case series can be collections of patients with a single outcome, looking at their antecedent exposures. For example, one might observe 100 consecutive women under the age of 50 who suffer from a pulmonary embolism, and note that 30 of them had been taking oral contraceptives.

After drug marketing, case series are most useful for two related purposes. First, they can be useful for quantifying the incidence of an adverse reaction. Second, they can be useful for being certain that any particular adverse effect of concern does not occur in a population which is larger than that studied prior to drug marketing. The so-called Phase IV postmarketing surveillance study of prazosin was conducted for the former reason, to quantitate the incidence of first-dose syncope from prazosin [32]. The Phase IV postmarketing surveillance study of cimetidine [33] was conducted for the latter reason. Metiamide was an H-2 blocker, which was withdrawn after marketing outside the US because it caused agranulocytosis. Since cimetidine is chemically related to metiamide, there was a concern that cimetidine too might cause agranulocytosis [32]. In both examples, the manufacturer asked its sales representatives to recruit physicians to participate in the study. Each participating physician then enrolled the next series of patients for whom the drug was prescribed.

In this type of study, one can be more certain that the patients are probably typical of those with the exposure or with the disease, depending on the focus of the study. However, in the absence of a control group, one cannot be certain which features in the description of the patients are unique to the exposure or outcome. As an example, one might have a case series from a particular hospital of 100 individuals with a certain disease, and note that all were men over the age of 60. This might lead one to conclude that this disease seems to be associated with being a man over the age of 60. However, it would be clear that this would be an incorrect conclusion once one noted that the hospital this case series was drawn from was a Veterans Administration hospital, where most patients are men over the age of 60. In the previous example of pulmonary embolism and oral contraceptives, 30% of the women with pulmonary embolism had been using oral contraceptives. However, this information is not sufficient to determine whether this is higher, the same as, or even lower than would have been expected. For this reason, case series are also not very useful in determining causation, but provide clinical descriptions of a disease or of patients who receive an exposure.

Analyses of Secular Trends

Analyses of secular trends, also called "ecologic studies," examine trends in an exposure that is a presumed cause and trends in a disease that is a presumed effect and test whether the trends coincide. These trends can be examined over time or across geographic boundaries. In other words, one could analyze data from a single region and examine how the trend changes over time, or one could analyze data from a single time period and compare how the data differ from region to region or country to country. Vital statistics are often used for these studies. As an example, one might look at sales data for oral contraceptives and compare them to death rates from venous thromboembolism, using recorded vital statistics. When such a study was actually performed, mortality rates from venous

thromboembolism were seen to increase in parallel with increasing oral contraceptive sales, but only in women of reproductive age, not in older women or in men of any age [34].

Analyses of secular trends are useful for rapidly providing evidence for or against a hypothesis. However, these studies lack data on individuals; they utilize only aggregated group data (e.g., annual sales data in a given geographic region in relation to annual cause-specific mortality in the same region). As such, they are unable to control for confounding variables. Thus, among exposures whose trends coincide with that of the disease, analyses of secular trends are unable to differentiate which factor is likely to be the true cause. For example, lung cancer mortality rates in the US have been increasing in women, such that lung cancer is now the leading cause of cancer mortality in women [35]. This is certainly consistent with the increasing rates of cigarette smoking observed in women until the mid-1960s [36], and so appears to be supportive of the association between cigarette smoking and lung cancer. However, it would also be consistent with an association between certain occupational exposures and lung cancer, as more women in the US are now working outside the home.

Case–Control Studies

Case–control studies compare cases with a disease to controls without the disease, looking for differences in antecedent exposures. As an example, one could select cases of young women with venous thromboembolism and compare them to controls without venous thromboembolism, looking for differences in antecedent oral contraceptive use. Several such studies have been performed, generally demonstrating a strong association between the use of oral contraceptives and venous thromboembolism [37].

Case–control studies can be particularly useful when one wants to study multiple possible causes of a single disease, as one can use the same cases and controls to examine any number of exposures as potential risk factors. This design is also particularly useful when one is studying a relatively rare disease, as it guarantees a sufficient number of cases with the disease. Using case–control studies, one can study rare diseases with markedly smaller sample sizes than those needed for cohort studies (see Chapter 4). For example, the classic study of diethylstilbestrol and clear cell vaginal adenocarcinoma required only 8 cases and 40 controls [31], rather than the many thousands of exposed subjects that would have been required for a cohort study of this question.

Case–control studies generally obtain their information on exposures retrospectively; that is, by recreating events that happened in the past. Information on past exposure to potential risk factors is generally obtained by abstracting medical records or by administering questionnaires or interviews. As such, case–control studies are subject to limitations in the validity of retrospectively collected exposure information. In addition, the proper selection of controls can be a challenging task and appropriate control selection can lead to a selection bias, which may lead to incorrect conclusions. Nevertheless, when case–control studies are done well, subsequent well-done cohort studies or randomized clinical trials, if any, will generally confirm their results. As such, the case–control design is a very useful approach for pharmacoepidemiologic studies.

Cohort Studies

Cohort studies identify subsets of a defined population and follow them over time, looking for differences in their outcome. Cohort studies generally are used to compare exposed patients to unexposed patients, although they can also be used to compare one exposure to another. For example, one could compare women of reproductive age who use oral contraceptives to users of other contraceptive methods, looking

for the differences in the frequency of venous thromboembolism. When such studies were performed, they in fact confirmed the relationship between oral contraceptives and thromboembolism, which had been noted using analyses of secular trends and case–control studies [38,39]. Cohort studies can be performed either prospectively, that is simultaneous with the events under study, or retrospectively, that is after the outcomes under study had already occurred, by recreating those past events using medical records, questionnaires, or interviews.

The major difference between cohort and case–control studies is the basis upon which patients are recruited into the study (see Figure 3.2). Patients are recruited into case–control studies based on the presence or absence of a disease, and their antecedent exposures are then studied. Patients are recruited into cohort studies based on the presence or absence of an exposure, and their subsequent disease course is then studied.

Cohort studies have the major advantage of being free of the major problem that plagues case–control studies: the difficult process of selecting an undiseased control group. In addition, prospective cohort studies are free of the problem of the questionable validity of retrospectively collected data. For these reasons, an association demonstrated by a cohort study is more likely to be a causal association than one demonstrated by a case–control study. Furthermore, cohort studies are particularly useful when one is studying multiple possible outcomes from a single exposure, especially a relatively uncommon exposure. Thus, they are especially useful in postmarketing drug surveillance studies, which are looking at any possible effect of a newly marketed drug. However, cohort studies can require extremely large sample sizes to study relatively uncommon outcomes (see Chapter 4). In addition, prospective cohort studies can require a prolonged time period to study delayed drug effects.

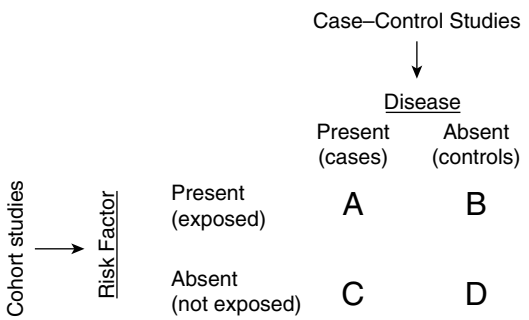


Figure 3.2 Cohort and case–control studies provide similar information, but approach data collection from opposite directions. *Source:* Strom BL. Medical databases in post-marketing drug surveillance. *Trends Pharmacol Sci* 1986; **7**: 377–80.

Analysis of Case–Control and Cohort Studies

As can be seen in Figure 3.2, both case–control and cohort studies are intended to provide the same basic information; the difference is how this information is collected. The key statistic reported from these studies is the *relative risk*, the ratio of the incidence rate of an outcome in the exposed group to the incidence rate of the outcome in the unexposed group. A relative risk of greater than 1.0 means that exposed subjects have a *greater* risk of the disease under study than unexposed subjects, or that the exposure appears to cause the disease. A relative risk of less than 1.0 means that exposed subjects have a *lower* risk of the disease than unexposed subjects, or that the exposure seems to protect against the disease. A relative risk of 1.0 means that exposed subjects and unexposed subjects have the same risk of developing the disease, or that the exposure and the disease appear unrelated.

One can calculate a relative risk directly from the results of a cohort study. However, in a case–control study one cannot determine the size of either the exposed population or the unexposed population from which the diseased cases and undiseased controls were drawn. The results of a case–control study do not provide

information on the incidence rates of the disease in exposed and unexposed individuals. Therefore, relative risks cannot be calculated directly from a case–control study. Instead, in reporting the results of a case–control study one generally reports the *odds ratio*, which is a close estimate of the relative risk when the disease under study is relatively rare. Since case–control studies are generally used to study rare diseases, there generally is very close agreement between the odds ratio and the relative risk, and the results from case–control studies are often loosely referred to as relative risks, although they are in fact odds ratios.

Both relative risks and odds ratios can be reported with *P values*. These P values allow one to determine if the relative risk is statistically significantly different from 1.0; that is, whether the differences between the two study groups are likely to be due to random variation or are likely to represent real associations.

Alternatively, and probably preferably, relative risks and odds ratios can be reported with *confidence intervals*, which are an indication of the range of relative risks within which the true relative risk for the entire theoretical population is most likely to lie. As an approximation, a 95% confidence interval around a relative risk means that we can be 95% confident that the true relative risk lies in the range between the lower and upper limits of this interval. If a 95% confidence interval around a relative risk excludes 1.0, then the finding is statistically significant with a P value of less than 0.05. A confidence interval provides much more information than a P value, however. As an example, a study that yields a relative risk (95% confidence interval) of 1.0 (0.9–1.1) is clearly showing that an association is very unlikely. A study that yields a relative risk (95% confidence interval) of 1.0 (0.1–100) provides little evidence for or against an association. Yet, both could be reported as a relative risk of 1.0 and a P value greater than 0.05. As another example, a study that yields a relative risk (95% confidence interval) of 10.0 (9.8–10.2) precisely quantifies a 10-fold increase in risk that is also statistically significant. A study that yields a relative risk (95% confidence interval) of 10.0 (1.1–100) says little, other than that an increased risk is likely. Yet, both could be reported as a relative risk of 10.0 ($P<0.05$). As a final example, a study yielding a relative risk (95% confidence interval) of 3.0 (0.98–5.0) is strongly suggestive of an association, whereas a study reporting a relative risk (95% confidence interval) of 3.0 (0.1–30) would not be. Yet, both could be reported as a relative risk of 3.0 ($P>0.05$).

Finally, another statistic that one can calculate from a cohort study is the excess risk, also called the risk difference or, sometimes, the attributable risk. Whereas the relative risk is the ratio of the incidence rates in the exposed group versus the unexposed groups, the excess risk is the arithmetic difference between the incidence rates. The relative risk is more important in considering questions of causation. The excess risk is more important in considering the public health impact of an association, as it represents the increased rate of disease due to the exposure. For example, oral contraceptives are strongly associated with the development of myocardial infarction in young women [37]. However, the risk of myocardial infarction in nonsmoking women in their 20s is so low, that even a fivefold increase in that risk would still not be of public health importance. In contrast, women in their 40s are at higher risk, especially if they are cigarette smokers as well. Thus, oral contraceptives should not be as readily used in these women [37].

As with relative risks, excess risks cannot be calculated from case–control studies, as incidence rates are not available. As with the other statistics, P values can be calculated to determine whether the differences between the two study groups could have occurred just by chance. Confidence intervals can be calculated around excess risks as well, and would be interpreted analogously.

Randomized Clinical Trials

Finally, *experimental studies* are studies in which the investigator controls the therapy that is to be received by each participant. Generally, an investigator uses that control to randomly allocate patients between or among the study groups, performing a *randomized clinical trial*. For example, one could theoretically randomly allocate sexually active women to use either oral contraceptives or no contraceptives, examining whether they differ in their incidence of subsequent venous thromboembolism. The major strength of this approach is random assignment, which is the only way to make it likely that the study groups are comparable in potential confounding variables that are either unknown or unmeasurable. For this reason, associations demonstrated in randomized clinical trials are more likely to be causal associations than those demonstrated using one of the other study designs reviewed here.

However, even randomized clinical trials are not without their problems. The randomized clinical trial just outlined, allocating women to receive contraceptives or no contraceptives, demonstrates the major potential problems inherent in the use of this study design. It would obviously be impossible to perform, ethically and logistically. In addition, randomized clinical trials are expensive and artificial. Inasmuch as they have already been performed prior to marketing to demonstrate each drug's efficacy, they tend to be unnecessary after marketing. They are likely to be used in pharmacoepidemiologic studies mainly for supplementary studies of drug efficacy [40]. However, they remain the "gold standard" by which the other designs must be judged. Indeed, with the publication of the results from the Women's Health Initiative indicating that combination hormone replacement therapy causes an increased risk of myocardial infarction rather than a decreased risk [41–44], there has been increased concern about reliance solely on nonexperimental methods to study drug safety after marketing [45–47], and we are seeing the use of massive randomized clinical trials as part of postmarketing surveillance (see Chapter 32).

Discussion

Thus, a series of different study designs are available (Table 3.4), each with their respective advantages and disadvantages. Case reports, case series, analyses of secular trends, case–control studies, and cohort studies have been referred to collectively as *observational study designs* or *nonexperimental study designs*, in order to differentiate them from experimental studies. In nonexperimental study designs the investigator does not control the therapy, but simply observes and evaluates the results of ongoing medical care. Case reports, case series, and analyses of secular trends have also been referred to as *descriptive studies*. Case–control studies, cohort studies, and randomized clinical trials all have control groups, and have been referred to as *analytic studies*. The analytic study designs can be classified in two major ways: by how subjects are selected into the study and by how data are collected for the study (see Table 3.5). From the perspective of how subjects are recruited into the study, case–control studies can be contrasted with cohort studies. Specifically, case–control studies select subjects into the study based on the presence or absence of a disease, while cohort studies select subjects into the study based on the presence or absence of an exposure. From this perspective, randomized clinical trials can be viewed as a subset of cohort studies, a type of cohort study in which the investigator controls the allocation of treatment, rather than simply observing ongoing medical care. From the perspective of timing, data can be collected *prospectively*, that is simultaneously with the events under study, or *retrospectively*, that is after the events under

study had already developed. In the latter situation, one recreates events that happened in the past using medical records, questionnaires, or interviews. Data can also be collected using *cross-sectional studies*, studies that have no time sense, as they examine only one point in time. In principle, either cohort or case–control studies can be performed using any of these time frames, although prospective case–control studies are unusual. Randomized clinical trials must be prospective, as this is the only way an investigator can control the therapy received.

The terms presented in this chapter, which are those that will be used throughout the book, are probably those used by a majority of epidemiologists. Unfortunately, however, other terms have been used for most of these study designs as well. Table 3.5 presents several of the synonyms that have been used in the medical literature. The same term is sometimes used by different authors to describe different concepts. For example, in this book we are reserving the use of the terms "retrospective study" and "prospective study" to refer to a time sense. As is apparent from Table 3.5, however, in the past some authors have used the term "retrospective study" to refer to a case–control study and "prospective study" to refer to a cohort study, confusing the two concepts inherent in the classification schemes presented in the table. Other authors use the term "retrospective study" to refer to any nonexperimental study, while others appear to use it to refer to any study they do not like, as a term of derision! Unfortunately, when reading a scientific paper, there is no way of determining which usage the author intended. More important than the terminology, however, are the concepts underlying the terms. Once they understand these concepts, readers can choose to use whatever terminology they are comfortable with.

Table 3.5 Epidemiologic study designs.

A) Classified by how subjects are recruited into the study
1) Case–control (case-history, case-referent, retrospective, trohoc) studies
2) Cohort (follow-up, prospective) studies
3) Experimental studies (clinical trials, intervention study)
B) Classified by how data are collected for the study
1) Retrospective (historical, nonconcurrent, retrolective) studies
2) Prospective (prolective) studies
3) Cross-sectional studies

Conclusion

From the material presented in this chapter, it is hopefully now apparent that each study design has an appropriate role in scientific progress. In general, science proceeds from the bottom of Table 3.4 upward, from case reports and case series that are useful for suggesting an association to analyses of trends and case–control studies that are useful for exploring these associations. Finally, if a study question warrants the investment and can tolerate the delay until results become available, then cohort studies and randomized clinical trials can be undertaken to assess these associations more definitively.

For example, regarding the question of whether oral contraceptives cause venous thromboembolism, an association was first suggested by case reports and case series, then was explored in more detail by analyses of trends and a series of case–control studies [37]. Later, because of the importance of oral contraceptives, the number of women using them, and the fact that users were predominantly healthy women, the investment was made in two long-term, large-scale cohort studies [38,39]. This question might even be worth the investment of a randomized clinical trial, except it would not be feasible or ethical. In contrast, when thalidomide was marketed, it was

not a major breakthrough; other hypnotics were already available. Case reports of phocomelia in exposed patients were followed by case–control studies [48] and analyses of secular trends [49]. Inasmuch as the adverse effect was so terrible and the drug was not of unique importance, the drug was then withdrawn, without the delay that would have been necessary if cohort studies and/or randomized clinical trials had been awaited. Ultimately, a retrospective cohort study was performed, comparing those exposed during the critical time period to those exposed at other times [50].

In general, however, clinical, regulatory, commercial, and legal decisions need to be made based on the best evidence available at the time of the decision. To quote Sir Austin Bradford Hill [27]:

> All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time. Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8:30 next day.

References

1 Schneider D, Lilienfeld DE. *Lilienfeld's Foundations of Epidemiology*, 4th edn. New York: Oxford University Press, 2014.
2 MacMahon B. *Epidemiology: Principles and Methods*, 2nd edn. Hagerstown, MD: Lippincott-Raven, 1997.
3 Friedman G. *Primer of Epidemiology*, 5th edn. New York: McGraw-Hill, 2003.
4 Mausner JS, Kramer S. *Epidemiology: An Introductory Text*, 2nd edn. Philadelphia, PA: Saunders, 1985.
5 Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd edn. Boston, MA: Little Brown, 1991.
6 Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology, 3rd edn*. Philadelphia, PA: Lippincott Williams & Wilkins, 2012.
7 Weiss N. *Clinical Epidemiology: The Study of the Outcome of Illness*, 3rd edn. New York: Oxford University Press, 2006.
8 Kelsey JL, Whittemore AS, Evans AS. *Methods in Observational Epidemiology*, 2nd edn. New York: Oxford University Press, 1996.
9 Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, MA: Little Brown, 1987.
10 Fletcher RH, Fletcher SW, Fletcher GS. *Clinical Epidemiology: The Essentials, 5th edn*. Lippincott Williams & Wilkins, 2012.
11 Gordis L. *Epidemiology*, 5th edn. Philadelphia, PA: Saunders, 2013.
12 Rothman KJ. *Epidemiology: An Introduction*, 2nd edn. New York: Oxford University Press, 2012.
13 Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*, 3rd edn. Sudbury, MA: Jones and Bartlett, 2014.
14 Katz DL, Wild D, Elmore JG, Lucan SC. *Jekel's Epidemiology, Biostatistics, and Preventive Medicine*, 4th edn. Philadelphia, PA: Elsevier, 2014.
15 Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*, 3rd edn. Sudbury, MA: Jones and Bartlett, 2014.
16 Friis RH, Sellers TA. *Epidemiology for Public Health Practice*, 5th edn. Sudbury, MA: Jones & Bartlett, 2014.

17 Hulley SB, Cummings SR, Browner WS, Grady D, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*, 4th edn. Baltimore, MD: Lippincott Williams & Wilkins, 2013.

18 Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR. *Medical Epidemiology*, 4th edn. New York: McGraw-Hill, 2004.

19 Wassertheil-Smoller S, Smoller J. *Biostatistics and Epidemiology: A Primer for Health Professionals*, 4th edn. New York: Springer, 2015.

20 Katz DL. *Clinical Epidemiology and Evidence-Based Medicine: Fundamental Principles of Clinical Reasoning and Research*. Thousand Oaks, CA: Sage, 2001.

21 Bhopal RS. *Concepts of Epidemiology: Integrating the Ideas, Theories, Principles and Methods of Epidemiology*, 3rd edn. New York: Oxford University Press, 2016.

22 Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical Epidemiology: How to Do Clinical Practice Research*, 3rd edn. Philadelphia, PA: Lippincott-Raven, 2006.

23 Woodward M. *Epidemiology: Study Design and Data Analysis*, 3rd edn. Boca Raton, FL: Chapman & Hall/CRC, 2013.

24 Weiss NS, Koepsell TD. *Epidemiologic Methods: Studying the Occurrence of Illness*. New York: Oxford University Press, 2014.

25 Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; **32**: 51–63.

26 Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.

27 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**: 295–300.

28 US Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Government Printing Office, 1964, p. 20.

29 Experimental carcinogenesis with tobacco smoke. In: *US Public Health Service: The Health Consequences of Smoking – Cancer. A Report of the Surgeon General*. Washington, DC: Government Printing Office, 1982, p. 181.

30 Biomedical evidence for determining causality. In: *US Public Health Service: The Health Consequences of Smoking – Cancer. A Report of the Surgeon General*. Washington, DC: Government Printing Office, 1982, p. 13.

31 Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971; **284**: 878–81.

32 Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC: Joint Commission on Prescription Drug Use, 1980.

33 Humphries TJ, Myerson RM, Gifford LM, *et al*. A unique postmarket outpatient surveillance program of cimetidine: report on phase II and final summary. *Am J Gastroenterol* 1984; **79**: 593–6.

34 Markush RE, Seigel DG. Oral contraceptives and mortality trends from thromboembolism in the United States. *Am J Public Health* 1969; **59**: 418–34.

35 National Center for Health Statistics. *Health: United States 1982*. Hyattsville, MD: Department of Health and Human Services, 1982.

36 US Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Government Printing Office, 1964, p. A1.

37 Strom BL, Stolley PD. Vascular and cardiac risks of steroidal contraception. In: Sciarra JW ed., *Gynecology and Obstetrics*, vol. **6**. Hagerstown, MD: Harper and Row, 1989, pp. 1–17.

38 Royal College of General Practitioners. *Oral Contraceptives and Health*. London: Pitman, 1974, ch. 7.

39 Vessey MP, Doll R, Peto R, Johnson B, Wiggins P. A long-term follow-up study of women using different methods of contraception – an interim report. *J Biosoc Sci* 1976; **8**: 373–427.

40 Bell RL, Smith EO. Clinical trials in post-marketing surveillance of drugs. *Control Clin Trials* 1982; **3**: 61–8.

41 Manson JE, Hsia J, Johnson KC, *et al*. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; **349**: 523–34.

42 Herrington DM, Howard TD. From presumed benefit to potential harm: hormone therapy and heart disease. *N Engl J Med* 2003; **349**: 519–21.

43 John B. Hormone-replacement therapy and cardiovascular diseases. *N Engl J Med* 2003; **349**: 521–2.

44 Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; **288**: 321–33.

45 Col NF, Pauker SG. The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? *Ann Intern Med* 2003; **139**: 923–9.

46 Grimes DA, Lobo RA. Perspectives on the Women's Health Initiative trial of hormone replacement therapy. *Obstet Gynecol* 2002; **100**: 1344–53.

47 Whittemore AS, McGuire V. Observational studies and randomized trials of hormone replacement therapy: what can we learn from them? *Epidemiology* 2003; **14**: 8–9.

48 Mellin GW, Katzenstein M. The saga of thalidomide. *N Engl J Med* 1962; **267**: 1238–44.

49 Taussig HB. A study of the German outbreak of phocomelia. *JAMA* 1962; **180**: 1106–14.

50 Kajii T, Kida M, Takahashi K. The effect of thalidomide intake during 113 human pregnancies. *Teratology* 1973; **8**: 163–6.

4

Sample Size Considerations for Pharmacoepidemiologic Studies

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

Chapter 1 pointed out that between 500 and 3000 subjects are usually exposed to a drug prior to marketing, in order to be 95% certain of detecting adverse effects that occur in between 1 and 6 in 1000 exposed individuals. While this seems like a reasonable goal, it poses some important problems that must be taken into account when planning pharmacoepidemiologic studies. Specifically, such studies must generally include a sufficient number of subjects to add significantly to the premarketing experience, and this requirement for large sample sizes raises logistical obstacles to cost-effective studies. This central special need for large sample sizes is what has led to the innovative approaches to collecting pharmacoepidemiologic data that are described in Part III of this book.

The approach to considering the implications of a study's sample size is somewhat different depending on whether a study is already completed or is being planned. After a study is completed, if a real finding was statistically significant, then the study had a sufficient sample size to detect it, by definition. If a finding was not statistically significant, then one can use either of two approaches. First, one can examine the resulting confidence intervals in order to determine the smallest differences between the two study groups that the study had sufficient sample size to exclude [1]. Alternatively, one can approach the question in a manner similar to the way one would approach it if one were planning the study *de novo*. Nomograms can be used to assist a reader in interpreting negative clinical trials in this way [2].

In contrast, in this chapter we will discuss in more detail how to determine a proper study sample size, from the perspective of one who is designing a study *de novo*. Specifically, we will begin by discussing how one calculates the minimum sample size necessary for a pharmacoepidemiologic study, to avoid the problem of a study with a sample size that is too small. We will first present the approach for cohort studies, then for case–control studies, and then for case series. For each design, one or more tables will be presented to assist the reader in carrying out these calculations.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Sample Size Calculations for Cohort Studies

The sample size required for a cohort study depends on what you are expecting from the study. To calculate sample sizes for a cohort study, one needs to specify five variables (see Table 4.1) [3,4].

The first variable to specify is the *alpha* (α) *or type I error* that one is willing to tolerate in the study. Type I error is the probability of concluding that there is a difference between the groups being compared when in fact a difference does not exist. Using diagnostic tests as an analogy, a type I error is a false positive study finding. The more tolerant one is willing to be of type I error, the smaller the sample size required. The less tolerant one is willing to be of type I error, the smaller one would set alpha, and the larger the sample size that would be required. Conventionally alpha is set at 0.05, although this certainly does not have to be the case. Note that alpha needs to be specified as either one-tailed or two-tailed. If only one of the study groups could conceivably be more likely to develop the disease and one is interested in detecting this result only, then one would specify alpha to be one-tailed. If either of the study groups may be likely to develop the disease and either result would be of interest, then one would specify alpha to be two-tailed. To decide whether alpha should be one-tailed or two-tailed, investigators should consider what their reaction would be to a result that is statistically significant in a direction opposite to the one expected. For example, what if one observed that a drug increased the frequency of dying from coronary artery disease instead of decreasing it, as expected? If the investigator's response to this would be "Boy, what a surprise, but I believe it," then a two-tailed test should be performed. If the investigator's response would be "I don't believe it, and I will interpret this simply as a study that does not show the expected decrease in coronary artery disease in the group treated with the study drug," then a one-tailed test should be performed. The more conservative option is the two-tailed test, assuming that the results could turn out in either direction. This is the option that is usually, although not always, chosen.

Table 4.1 Information needed to calculate a study's sample size.

For cohort studies	For case–control studies
1) Alpha, or type I error, considered tolerable, and whether it is one-tailed or two-tailed	1) Alpha, or type I error, considered tolerable, and whether it is one-tailed or two-tailed
2) Beta, or type II error, considered tolerable	2) Beta, or type II error, considered tolerable
3) Minimum relative risk to be detected	3) Minimum relative risk to be detected
4) Incidence of the disease in the unexposed control group	4) Prevalence of the exposure in the undiseased control group
5) Ratio of unexposed controls to exposed study subjects	5) Ratio of undiseased controls to diseased study subjects

The second variable that needs to be specified to calculate a sample size for a cohort study is the *beta* (β) *or type II error* that one is willing to tolerate in the study. A type II error is the probability of concluding that there is no difference between the groups being compared when in fact a difference does exist. In other words, a type II error is the probability of missing a real difference. Using diagnostic tests as an analogy, a type II error is a false negative study finding. The complement of beta is the power of a study; that is, the probability of detecting a difference if a difference really exists. Power is calculated as 1-β. Again, the more tolerant one is willing to be of type II errors – that is, the higher the beta – the smaller the sample size required. The beta is conventionally set at 0.1 (i.e., 90% power) or 0.2 (i.e., 80% power), although again this need not be the case. Beta is always one-tailed.

The third variable one needs to specify in order to calculate sample sizes for a cohort study is the minimum effect size one wants to be able to detect. For a cohort study, this is expressed as a relative risk. The smaller the relative risk that one wants to detect, the larger the sample size required. Note that the relative risk often used by investigators in this calculation is the relative risk the investigator is expecting from the study. This is *not correct*, as it will lead to inadequate power to detect relative risks that are smaller than expected, but still clinically important to the investigator. In other words, if one chooses a sample size that is designed to detect a relative risk of 2.5, one should be comfortable with the thought that, if the actual relative risk turns out to be 2.2, one may not be able to detect it as a statistically significant finding.

In a cohort study one selects subjects based on the presence or absence of an exposure of interest and then investigates the incidence of the disease of interest in each of the study groups. Therefore, the fourth variable one needs to specify is the expected incidence of the study outcome in the unexposed control group. Again, the more one asks of a study (e.g., the power to detect very small differences), the larger the sample size needed. Specifically, the rarer the outcome of interest, the larger the sample size needed.

The fifth variable one needs to specify is the number of unexposed control subjects to be included in the study for each exposed study subject. A study has the most statistical power for a given number of study subjects if it has the same number of exposed and unexposed subjects (controls). However, sometimes the number of exposed subjects is limited and, therefore, inadequate to provide sufficient power to detect a relative risk of interest. In that case, additional power can be gained by increasing the number of controls alone. Doubling the number of controls – that is, including two controls for each exposed subject – results in a modest increase in the statistical power, but it does not double it. Including three controls for each exposed subject increases the power further. However, the increment in power achieved by increasing the ratio of control subjects to exposed subjects from 2:1 to 3:1 is smaller than the increment in power achieved by increasing the ratio from 1:1 to 2:1. Each additional increase in the size of the control group increases the power of the study further, but with progressively smaller gains in statistical power. Thus, there is rarely a reason to include more than three or four controls per study subject. For example, one could design a study with an alpha of 0.05 to detect a relative risk of 2.0 for an outcome variable that occurs in the control group with an incidence rate of 0.01. A study with 2319 exposed individuals and 2319 controls would yield a power of 0.80, or an 80% chance of detecting a difference of that magnitude. With the same 2319 exposed subjects, ratios of control subjects to exposed subjects of 1:1, 2:1, 3:1, 4:1, 5:1, 10:1, and 50:1 would result in statistical powers of 0.80, 0.887, 0.913, 0.926, 0.933, 0.947, and 0.956, respectively.

It is important to differentiate between the number of controls (as has been discussed and illustrated) and the number of control groups. It is not uncommon, especially in case–control studies where the selection of a proper control group can be difficult, to choose more than one control group (for example, a group of hospital controls and a group of community controls). This is done for reasons of validity, not for statistical power, and it is important that these multiple control groups not be aggregated in the analysis. In this situation, the goal is to insure that the comparison of the exposed subjects to each of the different control groups yields the same answer, not to increase the available sample size. As such, the comparison of each control group to the exposed subjects should be treated as a separate study. The comparison of the exposed group to each control group requires a separate sample size calculation.

Once the five variables outlined have been specified, the sample size needed for a given study can be calculated. Several different formulas have been used for this calculation, each of which gives slightly different results. The formula that is probably most often used is modified from Schlesselman [3]:

$$N = \frac{1}{\left[p(1-R)^2\right]}\left[Z_{1-\alpha}\sqrt{\left(1+\frac{1}{K}\right)U(1-U)} + Z_{1-\beta}\sqrt{pR(1-Rp)+\frac{p(1-p)}{K}}\right]^2$$

where p is the incidence of the disease in the unexposed, R is the minimum relative risk to be detected, α is the type I error rate which is acceptable, β is the type II error rate which is acceptable, $Z_{1-\alpha}$ and $Z_{1-\beta}$ refer to the unit normal deviates corresponding to α and β, K is the ratio of number of unexposed control subjects to the number of exposed subjects, and

$$U = \frac{Kp+pR}{K+1}$$

$Z_{1-\alpha}$ is replaced by $Z_{1-\alpha/2}$ if one is planning to analyze the study using a two-tailed alpha. Note that K does not need to be an integer.

A series of tables are presented in Appendix A, which were calculated using this formula. In Tables A1–A4 we have assumed an alpha (two-tailed) of 0.05, a beta of 0.1 (90% power), and control to exposed ratios of 1:1, 2:1, 3:1, and 4:1, respectively. Tables A5–A8 are similar, except they assume a beta of 0.2 (80% power). Each table presents the number of exposed subjects needed to detect any of several specified relative risks, for outcome variables that occur at any of several specified incidence rates. The total study size will be the sum of exposed subjects (as listed in the relevant table) plus the controls.

For example, what if one wanted to investigate a new nonsteroidal anti-inflammatory drug that is about to be marketed, but premarketing data raised questions about possible hepatotoxicity? This would presumably be studied using a cohort study design and, depending upon the values chosen for alpha, beta, the incidence of the disease in the unexposed population, the relative risk one wants to be able to detect, and the ratio of control to exposed subjects, the sample sizes needed could differ markedly (see Table 4.2). For example, what if your goal was to study hepatitis that occurs, say, in 0.1% of all unexposed individuals? If one wanted to design a study with one control per exposed subject to detect a relative risk of 2.0 for this outcome variable, assuming an alpha (two-tailed) of 0.05 and a beta of 0.1, one could look in Table A1 and see that it would require 31 483 exposed subjects, as well as an equal number of unexposed controls. If one were less concerned with missing a real finding, even if it were there, one could change beta to 0.2, and the required sample size would drop to 23 518 (see Table 4.2 and Table A5). If one wanted to minimize the number of exposed subjects needed for the study, one could include up to four controls for each exposed subject (Table 4.2 and Table A8). This would result in a sample size of 13 402, with four times as many controls, a total of 67 010 subjects. Finally, if one considers it inconceivable that this new drug could *protect* against liver disease and one is not interested in that outcome, then one might use a one-tailed alpha, resulting in a somewhat lower sample size of 10 728, again with four times as many controls. Much smaller sample sizes are needed to detect relative risks of 4.0 or greater; these are also presented in Table 4.2.

In contrast, what if one's goal was to study elevated liver function tests, which, say, occur in 1% of an unexposed population? If one wants to detect a relative risk of 2 for this more common outcome variable, only 3104 subjects would be needed in each group, assuming a two-tailed alpha of 0.05, a beta of 0.1, and one control per

Table 4.2 Examples of sample sizes needed for a cohort study.

Hypothetical disease	Incidence rate assumed in unexposed	Alpha	Beta	Relative risk to be detected	Control : exposed ratio	Sample size needed in exposed group	Sample size needed in control group
Abnormal liver function tests	0.01	0.05 (2-tailed)	0.1	2	1	3104	3104
	0.01	0.05 (2-tailed)	0.2	2	1	2319	2319
	0.01	0.05 (2-tailed)	0.2	2	4	1323	5292
	0.01	0.05 (1-tailed)	0.2	2	4	1059	4236
	0.01	0.05 (2-tailed)	0.1	4	1	568	568
	0.01	0.05 (2-tailed)	0.2	4	1	425	425
	0.01	0.05 (2-tailed)	0.2	4	4	221	884
	0.01	0.05 (1-tailed)	0.2	4	4	179	716
Hepatitis	0.001	0.05 (2-tailed)	0.1	2	1	31 483	31 483
	0.001	0.05 (2-tailed)	0.2	2	1	23 518	23 518
	0.001	0.05 (2-tailed)	0.2	2	4	13 402	53 608
	0.001	0.05 (1-tailed)	0.2	2	4	10 728	42 912
	0.001	0.05 (2-tailed)	0.1	4	1	5823	5823
	0.001	0.05 (2-tailed)	0.2	4	1	4350	4350
	0.001	0.05 (2-tailed)	0.2	4	4	2253	9012
	0.001	0.05 (1-tailed)	0.2	4	4	1829	7316
Cholestatic jaundice	0.0001	0.05 (2-tailed)	0.1	2	1	315 268	315 268
	0.0001	0.05 (2-tailed)	0.2	2	1	235 500	235 500
	0.0001	0.05 (2-tailed)	0.2	2	4	134 194	536 776
	0.0001	0.05 (1-tailed)	0.2	2	4	107 418	429 672
	0.0001	0.05 (2-tailed)	0.1	4	1	58 376	58 376
	0.0001	0.05 (2-tailed)	0.2	4	1	43 606	43 606
	0.0001	0.05 (2-tailed)	0.2	4	4	22 572	90 288
	0.0001	0.05 (1-tailed)	0.2	4	4	18 331	73 324

exposed subject. Alternatively, if one wanted to detect the same relative risk for an outcome variable that occurred as infrequently as 0.0001, perhaps cholestatic jaundice, one would need 315 268 subjects in each study group.

Obviously, cohort studies can require very large sample sizes to study uncommon diseases. A study of uncommon diseases is often better performed using a case–control study design, as described in the previous chapter.

Sample Size Calculations for Case–Control Studies

The approach to calculating sample sizes for case–control studies is similar to the approach for cohort studies. Again, there are five variables that need to be specified, the values of which depend on what the investigator expects from the study (see Table 4.1). Three of these are alpha, or the type I error one is willing to tolerate; beta, or the type II error one is willing to tolerate; and the minimum odds ratio (an approximation of the relative risk) one wants to be able to detect. These are defined and described in the section on cohort studies.

In addition, in a case–control study one selects subjects based on the presence or absence of the disease of interest, and then investigates the prevalence of the exposure of interest in each study group. This is in contrast to a cohort study, in which one selects subjects based on the presence or absence of an exposure, and then studies whether or not the disease of interest develops in each group. Therefore, the fourth variable to be specified for a case–control study is the expected prevalence of the exposure in the undiseased control group, rather than the incidence of the disease of interest in the unexposed control group of a cohort study.

Finally, analogous to the consideration in cohort studies of the ratio of the number of unexposed control subjects to the number of exposed study subjects, one needs to consider in a case–control study the ratio of the number of undiseased control subjects to the number of diseased study subjects. The principles in deciding upon the appropriate ratio to use are similar in both study designs. Again, there is rarely a reason to include a ratio greater than 3:1 or 4:1. For example, if one were to design a study with a two-tailed alpha of 0.05 to detect a relative risk of 2.0 for an exposure which occurs in 5% of the undiseased control group, a study with 516 diseased individuals and 516 controls would yield a power of 0.80, or an 80% chance of detecting a difference of that size. Studies with the same 516 diseased subjects and ratios of controls to cases of 1:1, 2:1, 3:1, 4:1, 5:1, 10:1, and 50:1 would result in statistical powers of 0.80, 0.889, 0.916, 0.929, 0.936, 0.949, and 0.959, respectively.

The formula for calculating sample sizes for a case–control study is similar to that for cohort studies (modified from [3]):

$$N = \frac{1}{(p-V)^2}\left[Z_{1-\alpha}\sqrt{\left(1+\frac{1}{K}\right)U(1-U)} + Z_{1-\beta}\sqrt{p(1-p)/K + V(1-V)}\right]^2$$

where R, α, β, $Z_{1-\alpha}$, and $Z_{1-\beta}$ are as earlier, p is the prevalence of the exposure in the control group, and K is the ratio of undiseased control subjects to diseased cases,

$$U = \frac{p}{K+1}\left[K + \frac{R}{1+p(R-1)}\right]$$

and

$$V = \frac{pR}{1+p(R-1)}$$

Again, a series of tables that provide sample sizes for case–control studies is presented in Appendix A. In Tables A9–A12, we have assumed an alpha (two-tailed) of 0.05, a beta of 0.1 (90% power), and control to case ratios of 1:1, 2:1, 3:1, and 4:1, respectively. Tables A13–A16 are similar, except they assume a beta of 0.2 (80% power). Each table presents the number of diseased subjects needed to detect any of a number of specified relative risks, for a number of specified exposure rates.

For example, what if again one wanted to investigate a new nonsteroidal anti-inflammatory drug that is about to be marketed, but premarketing data raised questions about possible hepatotoxicity? This time, however, one is attempting to use a case–control study design.

Again, depending upon the values chosen of alpha, beta, and so on, the sample sizes needed could differ markedly (see Table 4.3). For example, what if one wanted to design a study with one control (undiseased subject) per diseased subject, assuming an alpha (two-tailed) of 0.05 and a beta of 0.1? The sample size needed to detect a relative risk of 2.0 for any disease would vary, depending upon the prevalence of use of the drug being studied. If one optimistically assumed that the drug will be used nearly as commonly as ibuprofen, by perhaps 1% of the population, then one could look in Table A9 and see that it would require 3210 diseased subjects and an equal number of undiseased controls. If one were less concerned with missing a real association, even if it existed, one could opt for a beta of 0.2, and the required sample size would drop to 2398 (see Table 4.3 and Table A13). If one wanted to minimize the number of diseased subjects needed for the study, one could include up to four controls for each diseased subject (Table 4.3 and Table A16). This would result in a sample size of 1370, with four times as many controls. Finally, if one considered it inconceivable that this new drug could *protect* against liver disease, then one might use a one-tailed alpha, resulting in a somewhat lower sample size of 1096, again with four times as many controls. Much smaller sample sizes are needed to detect relative risks of 4.0 or greater and are also presented in Table 4.3.

In contrast, what if one's estimates of the new drug's sales were more conservative? If one wanted to detect a relative risk of 2.0 assuming sales to 0.1% of the population, perhaps similar to tolmetin, then 31 588 subjects would be needed in each group, assuming a two-tailed alpha of 0.05, a beta of 0.1, and one control per diseased subject. In contrast, if one estimated the drug would be used in only 0.01% of the population (i.e., in controls without the study disease of interest), perhaps like phenylbutazone, one would need 315 373 subjects in each study group.

Obviously, case–control studies can require very large sample sizes to study relatively uncommonly used drugs. In addition, each disease of interest requires a separate case group and, thereby, a separate study. As such, as described in Chapter 3, studies of uncommonly used drugs and newly marketed drugs are usually better done using cohort study designs, whereas studies of rare diseases are better done using case–control designs.

Sample Size Calculations for Case Series

As described in Chapter 3, the utility of case series in pharmacoepidemiology is limited, as the absence of a control group makes causal inference difficult. Despite this, however, this is a design that has been used repeatedly. There are scientific questions that can be addressed using this design, and the collection of a control group equivalent in size to the case series would add considerable cost to the study. Case series are usually used in pharmacoepidemiology to quantitate better the incidence of a particular disease in patients exposed to a newly marketed drug. For example, in the "Phase IV" postmarketing drug surveillance study conducted for prazosin, the investigators collected a case series of 10 000 newly exposed subjects recruited through the manufacturer's sales force, to quantitate better the incidence of first-dose syncope, which was a well-recognized adverse effect of this drug [5,6]. Case series are normally used to determine whether a disease occurs more frequently than some predetermined incidence in exposed patients. Most often, the predetermined incidence of interest is zero, and one is looking for any occurrences of an extremely rare illness. As another example, when cimetidine was first marketed, there was concern over whether it could cause agranulocytosis, since it was closely related chemically to metiamide,

Table 4.3 Examples of sample sizes needed for a case–control study.

Hypothetical drug	Prevalence rate assumed in undiseased	Alpha	Beta	Odds ratio to be detected	Control : case ratio	Sample size needed in case group	Sample size needed in control group
Ibuprofen	0.01	0.05 (2-tailed)	0.1	2	1	3210	3210
	0.01	0.05 (2-tailed)	0.2	2	1	2398	2398
	0.01	0.05 (2-tailed)	0.2	2	4	1370	5480
	0.01	0.05 (1-tailed)	0.2	2	4	1096	4384
	0.01	0.05 (2-tailed)	0.1	4	1	601	601
	0.01	0.05 (2-tailed)	0.2	4	1	449	449
	0.01	0.05 (2-tailed)	0.2	4	4	234	936
	0.01	0.05 (1-tailed)	0.2	4	4	190	760
Tolmetin	0.001	0.05 (2-tailed)	0.1	2	1	31 588	31 588
	0.001	0.05 (2-tailed)	0.2	2	1	23 596	23 596
	0.001	0.05 (2-tailed)	0.2	2	4	13 449	53 796
	0.001	0.05 (1-tailed)	0.2	2	4	10 765	43 060
	0.001	0.05 (2-tailed)	0.1	4	1	5856	5856
	0.001	0.05 (2-tailed)	0.2	4	1	4375	4375
	0.001	0.05 (2-tailed)	0.2	4	4	2266	9064
	0.001	0.05 (1-tailed)	0.2	4	4	1840	7360
Phenylbutazone	0.0001	0.05 (2-tailed)	0.1	2	1	315 373	315 373
	0.0001	0.05 (2-tailed)	0.2	2	1	235 579	235 579
	0.0001	0.05 (2-tailed)	0.2	2	4	134 240	536 960
	0.0001	0.05 (1-tailed)	0.2	2	4	107 455	429 820
	0.0001	0.05 (2-tailed)	0.1	4	1	58 409	58 409
	0.0001	0.05 (2-tailed)	0.2	4	1	43 631	43 631
	0.0001	0.05 (2-tailed)	0.2	4	4	22 585	90 340
	0.0001	0.05 (1-tailed)	0.2	4	4	18 342	73 368

another H-2 blocker, which had been removed from the market in Europe because it caused agranulocytosis. This study also collected 10 000 subjects. It found only two cases of neutropenia, one in a patient who was also receiving chemotherapy. There were no cases of agranulocytosis [7].

To establish drug safety, a study must include a sufficient number of subjects to detect an elevated incidence of a disease, if it exists. Generally, this is calculated by assuming that the frequency of the event in question is vanishingly small, so that the occurrence of the event follows a Poisson distribution, and then one generally calculates 95% confidence intervals around the observed results.

Table A17 in the Appendix is useful for making this calculation [8]. In order to apply this table, one first calculates the incidence rate observed from the study's results; that is, the

number of subjects who develop the disease of interest during the specified time interval, divided by the total number of individuals in the population at risk. For example, if three cases of liver disease were observed in a population of 1000 patients exposed to a new nonsteroidal anti-inflammatory drug during a specified period of time, the incidence would be 0.003. The number of subjects who develop the disease is the "Observed number on which estimate is based (n)" in Table A17. In this example, it is 3. The lower boundary of the 95% confidence interval for the incidence rate is then the corresponding "Lower limit factor (L)" multiplied by the observed incidence rate. In this example, it would be 0.206 × 0.003 = 0.000618. Analogously, the upper boundary would be the product of the corresponding "Upper limit factor (U)" multiplied by the observed incidence rate. In the example, this would be 2.92 × 0.003 = 0.00876. In other words, the incidence rate (95% confidence interval) would be 0.003 (0.000618–0.00876). Thus, the best estimate of the incidence rate would be 30 per 10 000, but there is a 95% chance that it lies between 6.18 per 10 000 and 87.6 per 10 000.

In addition, a helpful simple guide is the so-called rule of threes, useful in the common situation where no events of a particular kind are observed [8]. Specifically, if no events of a particular type (i.e., the events of interest to the study) are observed in a study of X individuals, then one can be 95% certain that the event occurs no more often than 3/X. For example, if 500 patients are studied prior to marketing a drug, then one can be 95% certain that any event which does not occur in any of those patients may occur with a frequency of 3 or less in 500 exposed subjects, or that it has an incidence rate of less than 0.006. If 3000 subjects are exposed prior to drug marketing, then one can be 95% certain that any event which does not occur in this population may occur in no more than 3 in 3000 subjects, or the event has an incidence rate of less than 0.001. Finally, if 10 000 subjects are studied in a postmarketing drug surveillance study, then one can be 95% certain that any events which are not observed may occur in no more than 3 in 10 000 exposed individuals, or that they have an incidence rate of less than 0.0003. In other words, events not detected in the study may occur less often than in 1 in 3333 subjects in the general population.

Discussion

The discussions about sample size determinations in cohort and case–control studies assume that one is able to obtain information on each of the five variables that factor into these sample size calculations. Is this realistic? Four of the variables are, in fact, totally in the control of the investigators, subject to their specification: alpha, beta, the ratio of control subjects to study subjects, and the minimum relative risk to be detected. Only one of the variables requires data derived from other sources. For cohort studies, this is the expected incidence of the disease in the unexposed control group. For case–control studies, it is the expected prevalence of the exposure in the undiseased control group. In considering this needed information, it is important to realize that the entire process of sample size calculation is approximate, despite its mathematical sophistication. There is certainly no compelling reason why an alpha should be 0.05, as opposed to 0.06 or 0.04. The other variables specified by the investigators are similarly arbitrary. As such, only an approximate estimate is needed for this missing variable. Often the needed information is readily available from some existing data source, for example vital statistics or commercial drug utilization data sources. If not, one can search the medical literature for one or more studies that have collected these data for a defined population, either deliberately or as a by-product of their data-collecting effort, and assume that the population one will study will be similar. If this is not an

appropriate assumption, or if no such data exist in the medical literature, one is left with two alternatives. The first, and better, alternative is to conduct a small pilot study within one's population, in order to measure the information one needs. The second is simply to guess. In the second case, one should consider what a reasonable higher guess and a reasonable lower guess might be as well, to see if the sample size should be increased to take into account the imprecision of the estimate.

Finally, what if one is studying multiple outcome variables (in a cohort study) or multiple exposure variables (in a case–control study), each of which differs in the frequency one expects in the control group? In that situation, an investigator might base the study's sample size on the variable that leads to the largest requirement, and note that the study will have even more power for the other outcome (or exposure) variables. Regardless, it is usually better to have a somewhat larger sample size than the minimum, to allow some leeway if any of the underlying assumptions were wrong. This also will permit subgroup analyses with adequate power. In fact, if there are important subgroup analyses that represent *a priori* hypotheses that one wants to be able to evaluate, one should perform separate sample size calculations for those subgroups. In this situation, one should use the incidence of disease or prevalence of exposure that occurs in the subgroups, not that which occurs in the general population.

Note that sample size calculation is often an iterative process. There is nothing wrong with performing an initial calculation, realizing that it generates an unrealistic sample size, and then modifying the underlying assumptions accordingly. What is important is that investigators examine their final assumptions closely, asking whether, given the compromises made, the study is still worth undertaking.

Note also that this discussion was restricted to sample size calculations for dichotomous variables; that is, variables with only two options: a study subject either has a disease or does not have a disease. Information was not presented on sample size calculations for continuous outcome variables; that is, variables that have some measurement, such as height, weight, blood pressure, or serum cholesterol. Overall, the use of a continuous variable as an outcome variable, unless the measurement is extremely imprecise, will result in a marked increase in the power of a study. Details about this are omitted because epidemiologic studies unfortunately do not usually have the luxury of using such variables. Readers who are interested in more information on this can consult a textbook of sample size calculations [9].

All of the previous discussions have focused on calculating a minimum necessary sample size. This is the usual concern. However, two other issues specific to pharmacoepidemiology are important to consider as well. First, one of the main advantages of postmarketing pharmacoepidemiologic studies is the increased sensitivity to rare adverse reactions that can be achieved, by including a sample size larger than that used prior to marketing. Since between 500 and 3000 patients are usually studied before marketing, most pharmacoepidemiologic cohort studies are designed to include at least 10 000 exposed subjects. The total population from which these 10 000 exposed subjects would be recruited would need to be very much larger, of course. Case–control studies can be much smaller, but generally need to recruit cases and controls from a source population of equivalent size as for cohort studies. These are not completely arbitrary figures, but are based on the principles described in this chapter, applied to the questions which remain of great importance to address in a postmarketing setting. Nevertheless, these figures should not be rigidly accepted, but should be reconsidered for each specific study. Some studies will require fewer subjects, many will require more. To accumulate these sample sizes while performing cost-effective studies, several special techniques have

been developed, which are described in Part III of this book.

Second, because of the development of these new techniques and the development of large electronic data systems (see Part IIIb), pharmacoepidemiologic studies have the potential for the relatively unusual problem of *too large* a sample size. It is even more important than usual, therefore, when interpreting the results of studies that use these data systems, to examine their findings, differentiating clearly between statistical significance and clinical significance. With a very large sample size, one can find statistically significant differences that are clinically trivial. In addition, it must be kept in mind that subtle findings, even if statistically and clinically important, could easily have been created by biases or confounders (see Chapter 3). Subtle findings should not be ignored, but should be interpreted with caution.

References

1 Makuch RW, Johnson MF. Some issues in the design and interpretation of "negative" clinical trials. *Arch Intern Med* 1986; **146**: 986–9.

2 Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negative clinical studies. *Ann Intern Med* 1983; **99**: 248–51.

3 Schlesselman JJ. Sample size requirements in cohort and case-control studies of disease. *Am J Epidemiol* 1974; **99**: 381–4.

4 Stolley PD, Strom BL. Sample size calculations for clinical pharmacology studies. *Clin Pharmacol Ther* 1986; **39**: 489–90.

5 Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first dose phenomenon. *BMJ* 1976; **2**: 1293–4.

6 Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC: Joint Commission on Prescription Drug Use, 1980.

7 Gifford LM, Aeugle ME, Myerson RM, Tannenbaum PJ. Cimetidine postmarket outpatient surveillance program. *JAMA* 1980; **243**: 1532–5.

8 Haenszel W, Loveland DB, Sirken MG. Lung cancer mortality as related to residence and smoking history. I. White Males. *J Natl Cancer Inst* 1962; **28**: 947–1001.

9 Cohen J. *Statistical Power Analysis for the Social Sciences*. New York: Academic Press, 1977.

5

When Should One Perform Pharmacoepidemiologic Studies?

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

As discussed in the previous chapters, pharmacoepidemiologic studies apply the techniques of epidemiology to the content area of clinical pharmacology. This chapter will review when pharmacoepidemiologic studies should be performed. It will begin with a discussion of the various reasons why one might perform pharmacoepidemiologic studies. Central to many of these is one's willingness to tolerate risk. Whether one's perspective is that of a manufacturer, regulator, academician, or clinician, one needs to consider the risk of adverse reactions that one considers tolerable. Thus, the chapter will continue with a discussion of the difference between safety and risk. It will conclude with a discussion of the determinants of one's tolerance of risk.

Reasons to Perform Pharmacoepidemiologic Studies

The decision to conduct a pharmacoepidemiologic study can be viewed as similar to the regulatory decision about whether to approve a drug for marketing or the clinical decision about whether to prescribe a drug. In each case, decision-making involves weighing the costs and risks of a therapy against its benefits.

The main costs of a pharmacoepidemiologic study are obviously the costs (monetary, effort, time) of conducting the study itself. These costs clearly will vary, depending on the questions posed and the approach chosen to answer them. Generally, the cost per patient in a postmarketing study, with the exception of postmarketing randomized clinical trials, is likely to be at least an order of magnitude less than the cost of a premarketing study. Other costs to consider are the opportunity costs of other research that might be left undone if this research is performed.

One risk of conducting a pharmacoepidemiologic study is the possibility that it could identify an adverse outcome as associated with the drug under investigation when in fact the drug does not cause this adverse outcome. Another risk is that it could provide false reassurances about a drug's safety. Both these risks can be minimized by appropriate study designs, skilled researchers, and appropriate and responsible interpretation of the results obtained.

The benefits of pharmacoepidemiologic studies could be conceptualized in four different categories: regulatory, marketing, clinical, and

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

legal (see Table 5.1). Each will be of importance to different organizations and individuals involved in deciding whether to initiate a study. Any given study will usually be performed for several of these reasons, which will be discussed in turn.

Regulatory

Perhaps the most obvious and compelling reason to perform a postmarketing pharmacoepidemiologic study is regulatory: a plan for a postmarketing pharmacoepidemiologic study is required before the drug will be approved for marketing. Requirements for postmarketing research have become progressively more frequent in recent years. For example, in the 1970s the US Food and Drug Administration (FDA) required postmarketing research at the time of approval for about one third of drugs, a requirement which increased to over 70% in the 1990s [1]. Many of these required studies have been randomized clinical trials, designed to clarify residual questions about a drug's efficacy. Others focus on questions of drug toxicity. Often it is unclear whether the pharmacoepidemiologic study was undertaken in response to a regulatory requirement or in response to merely a "suggestion" by the regulator, but the effect is essentially the same. Early examples of studies conducted to address regulatory questions include the "Phase IV" cohort studies performed of cimetidine [2] and prazosin [3], discussed in Chapters 1 and 3. Now that the FDA has the authority to require such studies, such requirements are becoming more common.

Sometimes a manufacturer may offer to perform a pharmacoepidemiologic study with the hope that the regulatory agency might thereby expedite drug approval. If the agency believed that any new serious problem would be detected rapidly and reliably after marketing, it could feel more comfortable about releasing the drug sooner. Although it is difficult to assess the impact of volunteered postmarketing studies on regulatory decisions, the very large economic impact of an earlier approval has motivated some manufacturers to initiate such studies. In addition, in recent years regulatory authorities have occasionally released a particularly important drug after essentially only Phase II testing,

Table 5.1 Reasons to perform pharmacoepidemiologic studies.

- A) Regulatory
  - 1) Required
  - 2) To obtain earlier approval for marketing
  - 3) As a response to question by regulatory agency
  - 4) To assist application for approval for marketing elsewhere
- B) Marketing
  - 1) To assist market penetration by documenting the safety of the drug
  - 2) To increase name recognition
  - 3) To assist in repositioning the drug
    - a) Different outcomes, e.g., quality of life and economic
    - b) Different types of patients, e.g., the elderly
    - c) New indications
    - d) Less restrictive labeling
  - 4) To protect the drug from accusations about adverse effects
- C) Legal
  - 1) In anticipation of future product liability litigation
- D) Clinical
  - 1) Hypothesis testing
    - a) Problem hypothesized on the basis of drug structure
    - b) Problem suspected on the basis of preclinical or premarketing human data
    - c) Problem suspected on the basis of spontaneous reports
    - d) Need to better quantitate the frequency of adverse reactions
  - 2) Hypothesis generating – need depends on:
    - a) whether it is a new chemical entity
    - b) the safety profile of the class
    - c) the relative safety of the drug within its class
    - d) the formulation
    - e) the disease to be treated, including:
      - i) its duration
      - ii) its prevalence
      - iii) its severity
      - iv) whether alternative therapies are available

with the understanding that additional data would be gathered during postmarketing testing. For example, zidovudine was released for marketing after only limited testing, and not until later were additional data gathered on both safety and efficacy, data which indicated, among other things, that the doses initially recommended were too large [4].

Some postmarketing studies of drugs arise in response to case reports of adverse reactions reported to the regulatory agency. One response to such a report might be to suggest a labeling change. Often a more appropriate response, clinically and commercially, would be to propose a pharmacoepidemiologic study. This study would explore whether this adverse event in fact occurs more often in those exposed to the drug than would have been expected in the absence of the drug and, if so, how large is the increased risk of the disease. As an example, a Medicaid database was used to study hypersensitivity reactions to tolmetin [5], following reports about this problem to the FDA's Spontaneous Reporting System [6].

Finally, drugs are obviously marketed at different times in different countries. A postmarketing pharmacoepidemiologic study conducted in a country which marketed a drug relatively early could be useful in demonstrating the safety of the drug to regulatory agencies in countries which have not yet permitted its marketing. This is becoming increasingly feasible, as both the industry and the field of pharmacoepidemiology are becoming more international, and regulators are collaborating more.

Marketing

As will be discussed later in this chapter, pharmacoepidemiologic studies are performed primarily to obtain the answers to clinical questions. However, it is clear that a major underlying reason for some pharmacoepidemiologic studies is the potential marketing impact of those answers. In fact, some companies make the marketing branch of the company responsible for pharmacoepidemiology, rather than the medical branch.

Because of the known limitations in the information available about the effects of a drug at the time of its initial marketing, many physicians are appropriately hesitant to prescribe a drug until a substantial amount of experience in its use has been gathered. A formal postmarketing surveillance study can speed that process, as well as clarify advantages or disadvantages a drug has compared to its competitors.

A pharmacoepidemiologic study can also be useful to improve product name recognition. The fact that a study is underway will often be known to prescribers, as will its results once it is publicly presented and published. This increased name recognition will presumably help sales. An increase in a product's name recognition is likely to result particularly from pharmacoepidemiologic studies that recruit subjects for the study via prescribers. However, while this technique can be useful in selected situations, it is extremely expensive and less likely to be productive of scientifically useful information than most other alternatives available. In particular, the conduct of a purely marketing exercise under the guise of a postmarketing surveillance study, not designed to collect useful scientific information, is to be condemned [7]. It is misleading and could endanger the performance of future scientifically useful studies, by resulting in prescribers who are disillusioned and, thereby, reluctant to participate in future studies.

Pharmacoepidemiologic studies can also be useful to reposition a drug that is already on the market; that is, to develop new markets for the drug. One could explore different types of outcomes resulting from the use of the drug for the approved indication, for example the impact of the drug on the cost of medical care (see Chapter 34) and on patients' quality of life (see Chapter 42). One could also explore the use of the drug for the approved indication in types of patients other than those included in premarketing studies, for example in children, in the

elderly, or in patients with multiple comorbidities and/or taking many concomitant medications. By exploring unintended beneficial effects, or even drug efficacy (see Chapter 33), one could obtain clues to and supporting information for new indications for drug use. Finally, whether because of questions about efficacy or questions about toxicity, drugs are sometimes approved for initial marketing with restrictive labeling. For example, bretylium was initially approved for marketing in the US only for the treatment of life-threatening arrhythmias. Approval for more widespread use requires additional data. These data can often be obtained from pharmacoepidemiologic studies.

Finally, and perhaps most importantly, pharmacoepidemiologic studies can be useful to protect the major investment made in developing and testing a new drug. When a question arises about a drug's toxicity, it often needs an immediate answer, or else the drug may lose market share or even be removed from the market. Immediate answers are often unavailable, unless the manufacturer had the foresight to perform pharmacoepidemiologic studies in anticipation of this problem. Sometimes these problems can be specifically foreseen and addressed. More commonly, they are not. However, the availability of an existing cohort of exposed patients and a control group will often allow a much more rapid answer than would have been possible if the study had to be conducted *de novo*. One example of this is provided by the experience of Pfizer Pharmaceuticals, when the question arose about whether piroxicam (Feldene®) was more likely to cause deaths in the elderly from gastrointestinal bleeding than the other nonsteroidal anti-inflammatory drugs. Although Pfizer did not fund studies in anticipation of such a question, it was fortunate that several pharmacoepidemiologic research groups had data available on this question because of other studies that they had performed [8]. McNeil was not as fortunate when questions were raised about anaphylactic reactions caused by zomepirac. If the data it eventually was able to have [9] had been available at the time of the crisis, they might not have removed the drug from the market. Later, Syntex recognized the potential benefit, and the risk, associated with the marketing of parenteral ketorolac, and chose to initiate a postmarketing surveillance cohort study at the time of the drug's launch [10–12]. Indeed, the drug was accused of multiple different adverse outcomes, and it was only the existence of this study, and its subsequently published results, that saved the drug in its major markets.

Legal

Postmarketing surveillance studies can theoretically be useful as legal prophylaxis, in anticipation of eventually having to defend against product liability suits (see Chapter 9). One often hears the phrase "What you don't know, won't hurt you." However, in pharmacoepidemiology this view is shortsighted and, in fact, very wrong. All drugs cause adverse effects; the regulatory decision to approve a drug and the clinical decision to prescribe a drug both depend on a judgment about the relative balance between the benefits of a drug and its risks. From a legal perspective, to win a product liability suit using a legal theory of negligence, a plaintiff must prove causation, damages, and negligence. A pharmaceutical manufacturer that is a defendant in such a suit cannot change whether its drug causes an adverse effect. If the drug does, this will presumably be detected at some point. The manufacturer also cannot change whether the plaintiff suffered legal damages from the adverse effect; that is, whether the plaintiff suffered a disability or incurred expenses resulting from a need for medical attention. However, even if the drug did cause the adverse outcome in question, a manufacturer certainly could document that it was performing state-of-the-art studies to attempt to detect whatever toxic effects the drug had. In addition, such studies could make easier the

defense of totally groundless suits, in which a drug is blamed for producing adverse reactions it does not cause.

Clinical

Hypothesis Testing

The major reason for most pharmacoepidemiologic studies is hypothesis testing. The hypotheses to be tested can be based on the structure or the chemical class of a drug. For example, the cimetidine study mentioned earlier [2] was conducted because cimetidine was chemically related to metiamide, which had been removed from the market in Europe because it caused agranulocytosis. Alternatively, hypotheses can also be based on premarketing or postmarketing animal or clinical findings. For example, the hypotheses can come from spontaneous reports of adverse events experienced by patients taking the drug in question. The tolmetin [5], piroxicam [8], zomepirac [9], and ketorolac [10–12] questions mentioned are all examples of this. Finally, an adverse effect may clearly be due to a drug, but a study may be needed to quantitate its frequency. An example would be the postmarketing surveillance study of prazosin, performed to quantitate the frequency of first-dose syncope [3]. Of course, the hypotheses to be tested can involve beneficial drug effects as well as harmful drug effects, subject to some important methodologic limitations (see Chapter 33).

Hypothesis Generating

Hypothesis-generating studies are intended to screen for previously unknown and unsuspected drug effects. In principle, all drugs could, and perhaps should, be subjected to such studies. However, some drugs may require these studies more than others. This has been the focus of a formal study, which surveyed experts in pharmacoepidemiology [13].

For example, it is generally agreed that new chemical entities are more in need of study than what are called "me too" drugs. This is because the lack of experience with related drugs makes it more likely that the new drug has possibly important, unsuspected effects.

The safety profile of the class of drugs should also be important to the decision about whether to conduct a formal screening postmarketing surveillance study for a new drug. Previous experience with other drugs in the same class can be a useful predictor of what the experience with the new drug in question is likely to be. For example, with the finding that troglitazone had an increased risk of liver disease [14], that became a concern as well with the later thiazolidinediones, pioglitazone and rosiglitazone [15]. Similarly, with the finding that rofecoxib was associated with myocardial infarction, that became a concern as well with celecoxib [16].

The relative safety of the drug within its class can also be helpful. A drug that has been studied in large numbers of patients before marketing and appears safe relative to other drugs within its class is less likely to need supplementary postmarketing surveillance studies. An extension of this approach, of course, is comparative effectiveness research (see Chapter 26).

The formulation of the drug can be considered a determinant of the need for formal screening pharmacoepidemiologic studies. A drug that will, because of its formulation, be used mainly in institutions, where there is close supervision, may be less likely to need such a study. When a drug is used under these conditions, any serious adverse effect is likely to be detected, even without any formal study.

The disease to be treated is an important determinant of whether a drug needs additional postmarketing surveillance studies. Drugs used to treat chronic illnesses are likely to be used for a long period of time. As such, it is important to know their long-term effects. This cannot be addressed adequately in the relatively brief time available for each premarketing study. Also, drugs used to treat common diseases are important to study, as many patients are likely to be exposed to them. Drugs used to treat mild or

self-limited diseases need careful study too, because serious toxicity is less acceptable. This is especially true for drugs used by healthy individuals, such as contraceptives. On the other hand, when one is using a drug to treat individuals who are very ill, one is more tolerant of toxicity, assuming the drug is efficacious.

Finally, it is important to know whether alternative therapies are available. If a new drug is not a major therapeutic advance, since it will be used to treat patients who would have been treated with the old drug, one needs to be more certain of its relative advantages and disadvantages. The presence of significant adverse effects, or the absence of beneficial effects, is less likely to be tolerated for a drug that does not represent a major therapeutic advance.

Safety versus Risk

Clinical pharmacologists are used to thinking about drug "safety": the statutory standard that must be met before a drug is approved for marketing in the US is that it needs to be proven to be "safe and effective under conditions of intended use." It is important, however, to differentiate safety from risk. Virtually nothing is without some risks. Even staying in bed is associated with a risk of acquiring bed sores! Certainly no drug is completely safe. Yet, the unfortunate misperception by the public persists that drugs mostly are and should be without any risk at all. Use of a "safe" drug, however, still carries some risk. It would be better to think in terms of *degrees of safety*. Specifically, a drug "is safe if its risks are judged to be acceptable" [17]. Measuring risk is an objective but probabilistic pursuit. A judgment about safety is a personal and/or social value judgment about the acceptability of that risk. Thus, assessing safety requires two extremely different kinds of activities: measuring risk and judging the acceptability of those risks [17]. The former is the focus of much of pharmacoepidemiology and most of this book. The latter is the focus of the following discussion. More detail is presented in Chapter 39.

Risk Tolerance

Whether or not to conduct a postmarketing surveillance pharmacoepidemiologic study also depends on one's willingness to tolerate risk. From a manufacturer's perspective, one can consider this risk in terms of the risk of a potential regulatory or legal problem that may arise. Whether one's perspective is that of a manufacturer, regulator, academician, or clinician, one needs to consider the risk of adverse reactions that one is willing to accept as tolerable. There are several factors that can affect one's willingness to tolerate the risk of adverse effects from drugs (see Table 5.2). Some of these factors are related to the adverse outcome being studied. Others are related to the exposure and the setting in which the adverse outcome occurs.

Features of the Adverse Outcome

The severity and reversibility of the adverse reaction in question are of paramount importance to its tolerability. An adverse reaction that is severe is much less tolerable than one that is mild, even at the same incidence. This is especially true for adverse reactions that result in permanent harm, for example birth defects or death.

Another critical factor that affects the tolerability of an adverse outcome is the frequency of the adverse outcome in those who are exposed. Notably, this is *not* a question of the relative risk of the disease due to the exposure, but a question of the excess risk attributed to the drug of interest (see Chapter 3). Use of tampons is extraordinarily strongly linked to toxic shock: prior studies have shown relative risks of between 10 and 20. However, toxic shock is sufficiently uncommon that even a 10- to 20-fold increase in the risk of the disease still contributes an

Table 5.2 Factors affecting the acceptability of risks.

A) Features of the adverse outcome
1) Severity
2) Reversibility
3) Frequency
4) "Dread disease"
5) Immediate versus delayed
6) Occurs in all people vs. just in sensitive people
7) Known with certainty or not
B) Characteristics of the exposure
1) Essential versus optional
2) Present vs. absent
3) Alternatives available
4) Risk assumed voluntarily
5) Drug use will be as intended vs. misuse is likely
C) Perceptions of the evaluator

extraordinarily small excess risk of toxic shock syndrome in those who use tampons [18].

In addition, the particular disease caused by the drug is important to one's tolerance of its risks. Certain diseases are considered by the public to be "dread diseases," those that generate more fear and emotion than others. Examples are AIDS and cancer. It is less likely that the risk of a drug will be considered acceptable if it causes one of these diseases.

Another relevant factor is whether the adverse outcome is immediate or delayed. Most individuals are less concerned about delayed risks than immediate risks. This is one of the factors that have probably slowed the success of antismoking efforts. In part this is a function of denial; delayed risks seem as if they may never occur. In addition, the economic concept of "discounting" plays a role here. An adverse event in the future is less bad than the same event today, and a beneficial effect today is better than the same beneficial effect in the future. Something else may occur between now and then that could make that delayed effect irrelevant or, at least, mitigate its impact. Thus, a delayed adverse event may be worth incurring if it can bring about beneficial effects today.

It is also important whether the adverse outcome is a type A reaction or a type B reaction. As described in Chapter 1, type A reactions are the result of an exaggerated but otherwise usual pharmacologic effect of a drug. Type A reactions tend to be common, but they are dose related, predictable, and less serious. In contrast, type B reactions are aberrant effects of a drug. Type B reactions tend to be uncommon, are not related to dose, and are potentially more serious. They may be due to hypersensitivity reactions, immunologic reactions, or some other idiosyncratic reaction to the drug. Regardless, type B reactions are the more difficult to predict or even detect. If one can predict an adverse effect, then one can attempt to prevent it. For example, in order to prevent aminophylline-induced arrhythmias and seizures, one can begin therapy at lower doses and follow serum levels carefully. For this reason, all other things being equal, type B reactions are usually considered less tolerable.

Finally, the acceptability of a risk also varies according to how well established it is. The same adverse effect is obviously less tolerable if one knows with certainty that it is caused by a drug than if it is only a remote possibility.

Characteristics of the Exposure

The acceptability of a risk is very different depending upon whether an exposure is essential or optional. Major adverse effects are much more acceptable when one is using a therapy that can save or prolong life, such as chemotherapy for malignancies. On the other hand, therapy for self-limited illnesses must have a low risk to be acceptable. Pharmaceutical products intended for use in healthy individuals, such as vaccines and contraceptives, must be exceedingly low in risk to be considered acceptable.

The acceptability of a risk is also dependent on whether the risk is from the presence of a treatment or its absence. One could conceptualize deaths from a disease that can be treated by

a drug that is not yet on the market as an adverse effect from the absence of treatment. For example, the six-year delay in introducing beta-blockers into the US market has been blamed for resulting in more deaths than all recent adverse drug reactions combined [19]. As a society, we are much more willing to accept risks of this type than risks from the use of a drug that has been marketed prematurely. Physicians are taught *primum non nocere* – first do no harm. This is somewhat analogous to our willingness to allow patients with terminal illnesses to die from these illnesses without intervention, while it would be considered unethical and probably illegal to perform euthanasia. In general, we are much more tolerant of sins of omission than sins of commission.

Whether any alternative treatments are available is another determinant of the acceptability of risks. If a drug is the only available treatment for a disease, particularly a serious disease, then greater risks will be considered acceptable. This was the reason zidovudine was allowed to be marketed for the treatment of AIDS, despite its toxicity and the limited testing that had been performed [4]. Analogously, studies of toxic shock syndrome associated with the use of tampons were of public health importance, despite the infrequency of the disease, because consumers could choose among other available tampons that were shown to carry different risks [18].

Whether a risk is assumed voluntarily is also important to its acceptability. We are willing to accept the risk of death in automobile accidents more than the much smaller risk of death in airline accidents, because we control and understand the former and accept the attendant risk voluntarily. Some people even accept the enormous risks of death from tobacco-related disease, but would object strongly to being given a drug that was a small fraction as toxic. In general, it is agreed that patients should be made aware of possibly toxic effects of drugs they are prescribed. When a risk is higher than it is with the usual therapeutic use of a drug, as with an invasive procedure or an investigational drug, one usually asks the patient for formal informed consent. The fact that fetuses cannot make voluntary choices about whether or not to take a drug contributes to the unacceptability of drug-induced birth defects.

Finally, from a societal perspective, one needs to be concerned about whether a drug will be and is used as intended or whether misuse is likely. Misuse, in and of itself, can represent a risk of the drug. For example, a drug is considered less acceptable if it is addicting and, so, is likely to be abused. In addition, the potential for overprescribing by physicians can decrease the acceptability of the drug. For example, in the controversy about birth defects from isotretinoin, there was no question that the drug was a powerful teratogen, and that it was a very effective therapy for serious cystic acne refractory to other treatments. There was no question either about its effectiveness for less severe acne. However, that effectiveness led to its widespread use, including in individuals who could have been treated with less toxic therapies, and a larger number of pregnancy exposures, abortions, and birth defects than otherwise would have occurred [20].

Perceptions of the Evaluator

Finally, much depends ultimately upon the perceptions of the individuals who are making the decision about whether a risk is acceptable. In the US, there have been more than a million deaths from traffic accidents over the past 30 years; tobacco-related diseases kill the equivalent of three jumbo jet loads every day; and 3000 children are born each year with embryopathy from their mothers' use of alcohol in pregnancy [21]. Yet, these deaths are accepted with little concern, while the uncommon risk of an airplane crash or being struck by lightning generates fear. The decision about whether to allow isotretinoin to remain on the market hinged on whether the efficacy of the drug for a small number of people

who had a disease which was disfiguring but not life threatening was worth the birth defects that would result in some other individuals. There is no way to remove this subjective component from the decision about the acceptability of risks. Indeed, much more research is needed to elucidate patients' preferences in these matters. However, this subjective component is part of what makes informed consent so important. Most people feel that the final subjective judgment about whether an individual should assume the risk of ingesting a drug should be made by that individual, after education by their physician. However, as an attempt to assist that judgment, it is useful to have some quantitative information about the risks inherent in some other activities. Some such information is presented in Table 5.3.

Table 5.3 Annual risks of death from some selected hazards.

Hazard	Annual death rate (per 100 000 exposed individuals)
Heart disease (US, 1985)	261.4
Sport parachuting	190
Cancer (US, 1985)	170.5
Cigarette smoking (age 35)	167
Hang gliding (UK)	150
Motorcycling (US)	100
Power boat racing (US)	80
Cerebrovascular disease (US, 1985)	51
Scuba diving (US)	42
Scuba diving (UK)	22
Influenza (UK)	20
Passenger in motor vehicle (US)	16.7
Suicide (US, 1985)	11.2
Homicide (US, 1985)	7.5
Cave exploration (US)	4.5
Oral contraceptive user (age 25–34)	4.3
Pedestrian (US)	3.8
Bicycling (US)	1.1
Tornados (US)	0.2
Lightning (US)	0.05

Source: Data derived from [21–23].

Conclusion

This chapter reviewed when pharmacoepidemiologic studies should be performed. After beginning with a discussion of the various reasons why one might perform pharmacoepidemiologic studies, it reviewed the difference between safety and risk. It concluded with a discussion of the determinants of one's tolerance of risk. Now that it is hopefully clear when one might want to perform a pharmacoepidemiologic study, the next part of the book will provide perspectives on pharmacoepidemiology from some of the different fields that use it.

References

1 Department of Health and Human Services. Office of Inspector General. Postmarketing Studies of Prescription Drugs. May 1996. OEI-03-94-00760, p. 11. http://oig.hhs.gov/oei/reports/oei-03-94-00760.pdf (accessed April 2019).

2 Humphries TJ, Myerson RM, Gifford LM, *et al*. A unique postmarket outpatient surveillance program of cimetidine: report on phase II and final summary. *Am J Gastroenterol* 1984; **79**: 593–6.

3 Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC: Joint Commission on Prescription Drug Use, 1980.

4 Young FE. The role of the FDA in the effort against AIDS. *Public Health Rep* 1988; **103**: 242–5.

5 Strom BL, Carson JL, Schinnar R, Sim E, Morse ML. The effect of indication on the risk of hypersensitivity reactions associated with tolmetin sodium vs. other nonsteroidal antiinflammatory drugs. *J Rheumatol* 1988; **15**: 695–9.

6 Rossi AC, Knapp DE. Tolmetin-induced anaphylactoid reactions. *N Engl J Med* 1982; **307**: 499–500.

7 Strom BL; members of the ASCPT Pharmacoepidemiology Section. Position paper on the use of purported postmarketing drug surveillance studies for promotional purposes. *Clin Pharmacol Ther* 1990; **48**: 598.

8 Bortnichak EA, Sachs RM. Piroxicam in recent epidemiologic studies. *Am J Med* 1986; **81**: 44–8.

9 Strom BL, Carson JL, Morse ML, West SL, Soper KA. The effect of indication on hypersensitivity reactions associated with zomepirac sodium and other nonsteroidal antiinflammatory drugs. *Arthritis Rheum* 1987; **30**: 1142–8.

10 Strom BL, Berlin JA, Kinman JL, *et al*. Parenteral ketorolac and risk of gastrointestinal and operative site bleeding: a postmarketing surveillance study. *JAMA* 1996; **275**: 376–82.

11 Feldman HI, Kinman JL, Berlin JA, *et al*. Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193–9.

12 Hennessy S, Kinman JL, Berlin JA, *et al*. Lack of hepatotoxic effects of parenteral ketorolac in the hospital setting. *Arch Intern Med* 1997; **157**: 2510–14.

13 Rogers AS, Porta M, Tilson HH. Guidelines for decision making in postmarketing surveillance of drugs. *J Clin Res Pharmacol* 1990; **4**: 241–51.

14 Diabetes Prevention Program. Design and methods for a clinical trial in the prevention of type 2 diabetes. *Diabetes Care.* 1999; **22**: 623–34.

15 Scheen AJ. Hepatotoxicity with thiazolidinediones: is it a class effect? *Drug Saf* 2001; **24**: 873–88.

16 Mukherjee D, Nissen SE, Topol EJ Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA.* 2001; **286**: 954–9.

17 Lowrance WW. *Of Acceptable Risk*. Los Altos, CA: William Kaufmann, 1976.

18 Stallones RA. A review of the epidemiologic studies of toxic shock syndrome. *Ann Intern Med* 1982; **96**: 917–20.

19 Binns TB. Therapeutic risks in perspective. *Lancet* 1987; **2**: 208–9.

20 Marwick C. FDA ponders approaches to curbing adverse effects of drug used against cystic acne. *JAMA* 1988; **259**: 3225.

21 Urquhart J, Heilmann K. *Risk Watch: The Odds of Life*. New York: Facts on File, 1984.

22 O'Brien B. *"What Are My Chances Doctor?" A Review of Clinical Risks*. London: Office of Health Economics, 1986.

23 Silverberg E, Lubera JA. Cancer statistics. *CA Cancer J Clin* 1988; **38**: 5–22.

Part II

The Role of Pharmacoepidemiology in Different Sectors

6

The Role of Pharmacoepidemiology in the Healthcare System and Academia

Joshua J. Gagne and Jerry Avorn

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Every year prescribers and patients have more medications at their disposal, each with its own efficacy, side effects, and cost. When a new drug is introduced, its benefit/risk relationship is often understood in only a preliminary way, as is its cost-effectiveness. This provides a limited perspective on how it ideally should be used. High-profile withdrawals of drugs for safety reasons, along with prominent warnings about widely used medications that remain on the market, have caused physicians, patients, and policymakers to become more aware of drug safety concerns. At the same time, healthcare systems all over the globe are struggling with how to provide the most appropriate care in the face of rising costs and increasingly tight fiscal constraints. Pharmacoepidemiology can serve as a key tool for helping to address all of these concerns. These issues are growing throughout the healthcare system, and particularly in academic medical centers.

Once a drug is approved for marketing, it enters a complex healthcare system in which its prescription, its use by patients, and its outcomes often go largely unassessed. Until recently, scant attention has been paid to systematic surveillance of these actions, except for the atypical settings of some integrated healthcare delivery systems. The prevailing view has been that after the US Food and Drug Administration (FDA) or comparable national authority approves a drug, it is used at the discretion of the clinician, with little formal follow-up of the appropriateness or consequences of such decisions. The problem is made more acute by the fact that many regulatory agencies purposely (and often by statute) do not base their approval decisions on a medication's clinical or economic value compared to similar products; often superiority over placebo is sufficient for a drug to be approved. In addition, it is generally no one's responsibility (other than the harried prescriber) to determine how faithfully patients are adhering to the prescribed regimen. Increasingly, more attention is being paid to assessing the outcomes of medication use on a population level, considering what its useful and harmful outcomes are when it is taken by hundreds, thousands, or even millions of patients rather than by single individuals in a clinical trial or in routine practice. It is now widely appreciated that some adverse events can be identified and their risk quantified only by observing a drug's use in large numbers of patients. The best perspective on the impact of a medication on the health of the public requires measuring those outcomes in the healthcare

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

system itself, rather than for one person at a time. It is here that the insights of pharmacoepidemiology are playing an increasingly central role.

Driven by the pressures just noted, this situation is evolving, with growing appreciation of several important problems, each of which can be informed by the methods and tools of pharmacoepidemiology: (i) medications that seem acceptably safe on approval may prove to have important risks which were underappreciated at the time of approval; (ii) in typical practice, clinicians often make prescribing decisions that do not reflect the best evidence base or guideline recommendations; (iii) even this evidence base is often thinner than it should be because head-to-head comparisons of drug effectiveness or safety – either trial based or observational – have not been done [1]; (iv) as a result, inadequate information is available to inform decisions about which drugs work best, or most cost-effectively, for specific indications; and (v) patients frequently fail to take their medications as directed.

Pharmacoepidemiology is the core discipline required for a rigorous understanding of each of these areas, and to guide the development and evaluation of programs to address them. Many of these topics are discussed in detail in the chapters that follow; this chapter provides an overview of how the field and its methods can contribute to these larger themes in medical care delivery and health services research, from the perspective of academia.

The Drug Approval Process

Each national healthcare system must grapple with the following inherent paradox of pharmacology: a new therapy must be evaluated for approval when the available data on its benefits and harms are still modest. Yet, waiting until "all the evidence is in" can pose its own public health threat, if this prevents an important new treatment from being used by patients who need it. Since any medication that is effective is bound to have some adverse effect in some organ system in some patients at some doses, any approval must by definition be based on a judgment that a drug's efficacy is "worth it" in light of the known risks of the treatment. However, the trials conducted by a given drug manufacturer to win approval are often powered statistically (see Chapter 4) to demonstrate success for that single product in achieving a prespecified therapeutic endpoint. Especially when this is demonstration of superiority over placebo, and/or when the required endpoint is reaching a surrogate outcome – for example, a change in a laboratory test such as hemoglobin A1c or low-density lipoprotein (LDL) cholesterol – the number of subjects required for these exercises and the duration of the studies are often inadequate to reveal important safety problems if they are present. This is exacerbated by the extensive exclusion criteria for study participation, a particular problem for high-risk populations such as the frail elderly, pregnant women, and children (see also Chapter 22) [2].

As a result, additional methods need to be applied even to preapproval data to aggregate adverse events from multiple study populations to provide the power needed to assess safety. Meta-analysis (see Chapter 36) of adverse effects data from multiple preapproval trials represents the first opportunity to use these tools to inform the appropriate use of medications. This makes it possible to combine findings from different, smaller studies – many of them conducted before the drug in question was approved – to produce evidence of potential harm for drugs such as rofecoxib (cardiovascular harm) [3], rosiglitazone (myocardial infarction) [4], or the selective serotonin reuptake inhibitors (SSRIs) used in children (suicidality) [5].

These shortcomings of premarketing trials are likely to become even more salient as

regulators move toward alternative drug approval processes. The European Medicines Agency recently completed a pilot project to explore approval via adaptive pathways, which are intended to provide earlier and progressive patient access to new drugs. The US FDA maintains several expedited regulatory pathways. Postapproval monitoring of drugs approved through these pathways is even more critical, because safety information is more likely to emerge in the postmarketing setting for drugs approved by these pathways compared to conventional pathways [6,7]. In 2016 the US Congress enacted the 21st Century Cure Act, which included, among other sections, provisions that modify the data required for FDA approval. In particular, the law promotes the use of biomarkers and surrogate measures to support the approval of new drugs as well as "real-world evidence" from observational data to support supplemental indications for existing products [8]. As these new provisions are implemented, pharmacoepidemiology will have an even greater role in generating the real-world evidence for supplemental indications [9], and will be increasingly relied upon to evaluate the impact on clinical outcomes, both beneficial and harmful, of new drugs approved on less rigorous evidence.

Prescribing Practices

Once a drug has entered the healthcare delivery system, a growing literature documents several areas in which prescribing falls short of existing knowledge. These issues can also be elucidated using the tools of pharmacoepidemiology. First, and often neglected, is the issue of *underprescribing*. Studies of many important chronic diseases such as hypertension, hypercholesterolemia, and diabetes reveal that many patients with these conditions have not been diagnosed by their physicians, and, when they have, they are often not prescribed an adequate regimen to control their risks, or even any regimen at all [10]. Even with a database that includes only drug utilization information, pharmacoepidemiology makes it possible to achieve a good first approximation of the problem of undertreatment by measuring the age- and gender-adjusted prevalence of use of medications to manage specific chronic conditions by a given clinician, or a given practice or health system (see Chapter 18). When patterns of use are combined with other research on prescriber characteristics and decision-making, it becomes possible to identify more clearly when and how prescribing falls short, insights which can then be used to shape programs to improve care (see later discussion) [11].

When medications are used, there is good evidence that clinicians frequently do not prescribe regimens that are optimal, based on the available clinical evidence, or prescribe medications that may interact with other drugs a patient takes, or choose more expensive drugs when comparable generic preparations would work as well and be much more affordable. Pharmacoepidemiology makes it possible to assess the distribution of drugs used for a given indication by clinician, practice, or system, even if only drug utilization datasets are available, though it is necessary to take into account whether a given prescriber is a specialist who may see in referral most of the refractory patients cared for by colleagues.

When diagnostic data are also available, a more sophisticated approach can also take into account contraindications and compelling indications related to specific drug choices, to refine the assessment of the appropriateness of prescribing in an entire healthcare system, or for individual clinicians (see Chapter 19). Numerous studies have documented these shortfalls in several domains of care. For example, one study assessed all hypertension-related medication use and diagnoses in one large state-funded program of medications for the

elderly. The availability of clinical information made it possible to determine how well the regimen of each patient conformed to the recommendations of the then-current guidelines of the Joint National Committee (JNC) on Prevention, Detection, and Treatment of High Blood Pressure. This study found that a substantial proportion of treated hypertensive patients were not receiving a regiment consistent with JNC guidelines [12]. Often, such suboptimal prescribing involved omissions of an indicated class (e.g., angiotensin-converting enzyme inhibitors in patients with diabetes mellitus), or use of a calcium channel blocker when a beta-blocker would have been more appropriate (e.g., in a hypertensive patient who has had a myocardial infarction). Another analysis reviewed all clinical encounters of patients who had filled prescriptions for clopidogrel and found that about half did not have any evidence of conditions (such as coronary artery stenting) for which the drug had an approved indication, or any other evidence-based reasons for its use [13].

Moving up an additional level in database detail, more sophisticated health records systems are becoming available each year that integrate pharmacy data with information from clinical laboratories, electronic health records, registries, and sources of patient-reported information to measure the adequacy of use of cholesterol-lowering agents, diabetes drugs, antihypertensives, and other drugs. This makes it possible to assess the effectiveness of prescribing outcomes for a given clinician (or practice or system), by measuring how well target metrics such as normotension or goal LDL cholesterol or hemoglobin A1c are being achieved. In all these analyses, pharmacoepidemiology makes it possible to evaluate the appropriateness of medication use in selected populations, even if it cannot with certainty determine whether a given prescription in a particular patient was the best choice.

Evaluation of Patients' Use of Drugs in the Healthcare System

Even when a medication is appropriately prescribed, patients may underuse it or use it in unsafe ways. Underuse of needed drugs by patients is one of the most common medication-related problems, and one that can be readily identified by pharmacoepidemiology (see also Chapter 38) [14]. Although it is less striking than obvious drug-induced adverse events, underuse is probably responsible for at least as much morbidity and mortality, if not more. To be fully understood, this requires the kind of denominator-grounded population orientation of a pharmacoepidemiologic perspective, which is still lacking in many healthcare systems [15]. The clinical trialist or the treating physician focuses on patients who are assigned to receive a drug in a study, or who are prescribed a drug in practice, respectively. But by expanding the view to the larger population of people of which those study subjects or patients make up a subsample, the pharmacoepidemiologist can also take into account all those people with a particular diagnosis who are *not* taking a given drug or drug class, perhaps because their clinician did not prescribe treatment, or because the patient did not have access to the medication, or had stopped treatment because of side effects.

The failure of a patient to fill a prescribed medication has been described using various terms, each with its own sociocultural baggage. (In fact, even the word "failure" is loaded in this way.) The word *compliance* has been criticized because it is seen as depicting a master–subservient relationship between doctor and patient, implying that a "noncompliant" patient is engaging in a kind of misbehavior. Many prefer the term *adherence*, which is more neutral. *Persistence* refers to the degree to which a patient sticks with a prescribed regimen over time. *Intelligent nonadherence* (or intelligent noncompliance) describes a situation in which a patient stops a therapy because it is producing

excessively burdensome side effects or failing to relieve symptoms effectively. Until around 1990, the field of adherence research was understudied, and most clinicians assumed that after they wrote a prescription, a patient filled it and took it more or less as directed. Once large-scale computerized pharmacy claims datasets and the methods of pharmacoepidemiology made it possible to readily measure the prescription-filling behavior of large numbers of people, it became clear that this simple assumption was often false (see Chapter 38) [16,17].

Datasets based on the complete paid-claims files of drug benefit programs (see Chapter 12) provided the first means of studying adherence in defined populations. Because such a claim is usually necessary before the pharmacy can be paid, and because insured patients are unlikely to pay out of pocket to fill prescriptions outside the system, such datasets provide an excellent record of what medications are actually dispensed [18]. When such datasets are analyzed, a grim fact emerges: averaging across studies, about half of all medications prescribed for the treatment of chronic conditions such as hypercholesterolemia, elevated blood pressure, osteoporosis, glaucoma, and so on are not taken [19]. This causes a massive and still underappreciated shortfall – at both the clinical and public health levels – in the benefit that these regimens could generate in preventing myocardial infarction, strokes, fractures, or visual loss, respectively [20]. The full magnitude of this problem is still not completely appreciated by clinicians or policymakers.

Because many assessments of underuse are based on pharmacy-generated data on filled prescriptions, it is sometimes difficult to know whether nonuse of an indicated drug was the result of a failure of the patient to fill a prescription, or the failure of the clinician to write it. Electronic prescribing makes it possible to define this problem more precisely. As bad as the problem of low refill rates is, these newer analyses have made it clear that the situation is even worse [21]. One large study found that a fourth of initial prescriptions written electronically were never picked up at the pharmacy [22]. As a result, the approximately 50% rate of nonadherence seen over time in pharmacoepidemiologic datasets based on filled prescriptions is a best-case scenario, as it does not even take into account the additional millions of regimens that are not initiated by the patient. Terminology has evolved to define the two aspects of this problem: *secondary nonadherence* refers to the failure by a patient to continue to use a medication already begun; *primary nonadherence* occurs when a clinician writes a prescription that the patient does not even fill once. The magnitude of both primary and secondary nonadherence is substantial, and varies by drug class as well as by country [23–26].

These findings about adherence have implications for other aspects of pharmacoepidemiologic studies. First, they raise important concerns about the validity of large databases (such as the UK Clinical Practice Research Database and other electronic health record databases) that define drug exposure in terms of what a clinician prescribed, as opposed to what the patient actually obtained from the pharmacy (see Chapters 13 and 37). Second, the very high rates of nonuse in typical practice settings cast doubt on randomized trial-based assumptions about the clinical benefit, public health impact, and cost-effectiveness of many regimens in widespread use. This issue points up the value of the "real-world" analyses performed by pharmacoepidemiologists using data from typical practice settings (see Chapters 11–14).

Many pharmacoepidemiologic studies have attempted to identify risk factors for poor adherence, with the goal of helping prescribers to spot proactively which patients are likely to be nonadherent [27]. Yet this literature has identified remarkably few such predictors. High drug cost has been one, especially in patients without adequate pharmacy benefit insurance. Such studies have also demonstrated that

insured patients prescribed a higher-cost medication adhere less well to their regimens than those prescribed a lower-cost generic in the same therapeutic class [28,29]. Several studies have identified claims-based measures of patients' prior adherence to other medications as a strong predictor of their future adherence to a new medication, likely reflecting their general willingness or ability to adhere to prescribed treatments [30,31]. Another consistent risk factor has been race, suggesting an important problem in physician–patient communication and/or trust for nonwhite patients [32]. However, other variables such as physician characteristics, or patient age, level of education, or morbidity, have not consistently been found to be associated with poor medication adherence, making the management of this common problem even more difficult.

Assessment of the Quality and Outcomes of Medication Use in Populations

Much attention is now being paid to assessment of the outcomes of medication use in typical real-world populations. This perspective is based on the difference between *efficacy,* the effect of a medication in the rigorous but idealized setting of a clinical trial, and its *effectiveness,* a measure of its outcomes in typical practice settings (see Chapter 33). These often differ. For example, one important conventional randomized trial demonstrated convincingly that the addition of spironolactone to the regimen of patients with congestive heart failure substantially improved their clinical status and reduced mortality [33]. However, a population-based analysis later established that when these findings were applied in routine practice by typical physicians treating a much larger number of typical patients, there was a significant increase in hyperkalemia-associated morbidity and mortality [34]. By contrast, an analysis of prescribers' responses to a different study that provided new evidence about the optimal management of atrial fibrillation demonstrated a more positive change in practice [35].

Other "lost in translation" analyses document that despite overwhelming randomized trial evidence showing the efficacy of warfarin use in preventing stroke in patients with atrial fibrillation, population-based studies of older patients living in nursing homes revealed a surprisingly low prevalence of use of this therapy [36]. Such underuse was found to be associated with physicians' recent experience with adverse events caused by the drug [37], as well as by their perceptions of and attitudes to risks and benefits [38]. Other pharmacoepidemiologic studies of medication use in nursing homes have documented similar dramatically low use of other well-documented medications in these high-risk populations, such as drugs to treat osteoporosis, even in patients who have already had a hip fracture [39]. This kind of real-world population research can lay the foundation for enlightened interventions to address such nonuse, by taking on its underlying causes.

Pharmacoepidemiologic methods can also be used to track the diffusion of new medication classes into practice [40], as well as the reaction of practitioners in various settings to new information about the comparative benefits and risks of drugs, as in the case of warnings about the cardiovascular toxicity of rosiglitazone [41]. Acknowledging the gap between the characteristics of clinical trial participants and those who often use a medication in practice, methods are also being developed and applied to generalize trial results to more typical patient populations. For example, the newer oral anticoagulant, dabigatran, was approved on the basis of a large randomized trial comparing it to warfarin, an older oral anticoagulant. A simulation-based approach was used to assess how the comparative benefits and risks would translate to cohorts of patients who use these drugs outside of the

randomized trial and in usual routine care [42]. Such an approach preserves the strengths of the randomized trial, but makes the results more useful to patients and clinicians.

Policy Analysis

Usually, policy changes are implemented in the healthcare system with no systematic plans for their evaluation, and no follow-up studies of their impact; this can be hyperbolically but poignantly characterized as a form of large-scale, sloppy human experimentation without informed consent. Such changes in benefit design are often applied to medication use. However, even if a policy is changed in a way that does not anticipate an evaluation, population-based observational studies after the fact can still yield important conclusions concerning its effects, both good and bad. For example, when the Canadian province of British Columbia implemented a reference-pricing policy for antihypertensive medications in which it reimbursed only the cost of an effective generic drug in several classes, critics charged that any savings would come at the cost of increased morbidity and healthcare utilization. However, a careful time-series analysis of all medication use, physician visits, and hospital care in the province before and after policy implementation provided compelling evidence that the new reimbursement system produced no important clinical downsides, but did achieve substantial savings for the provincial healthcare budget [43]. Such observational methods have also been combined with population-based randomized policy trials, and were found to yield similar results [43].

Similarly, one large US employer introduced a change in its drug benefit plan that reduced or eliminated patient co-payment requirements for cholesterol-lowering drugs and an expensive antiplatelet agent. While this new policy seemed intuitively appealing, no plan had been put in place to determine whether the additional costs incurred by the employer would result in patient benefit. A pharmacoepidemiologic analysis compared adherence rates to these medications by employees of that company with rates for comparable people insured by similar employers with less generous drug benefit plans, and found that the change in benefit design significantly improved adherence [44]. A similar policy was later tested by randomizing patients in a large health insurance plan who had recently been discharged from the hospital with myocardial infarction to either their usual drug benefit plan or to no co-payments for their postdischarge statins, beta-blockers, angiotensin-converting enzyme inhibitors, or angiotensin-receptor blockers [45]. The elimination of co-payments improved adherence without increasing total health spending, and also reduced the incidence of subsequent cardiac events. The tools of pharmacoepidemiology have also been used to evaluate other interventions to improve medication adherence, such as automated refill programs [46], automated and live reminder calls from pharmacists [47], and medication synchronization programs, which are designed to align patients' prescription refills on the same date to reduce the number of pharmacy visits per month [48].

Not all such policy interventions are as well conceived. Hard-pressed governmental programs such as Medicaid must often resort to prior approval requirements for certain costly drugs, which require prescribers to seek permission from the program before a given medication is dispensed. Sometimes the criteria that determine whether permission is granted are evidence based and plausible; other times they are not [49,50]. The methods of pharmacoepidemiology are increasingly used to assess the clinical and economic consequences of such policies [51–53]. One study documented an increase in the use of clopidogrel in one Canadian province after a highly restrictive policy was replaced with a more lenient one; this

change was associated with a significant concomitant reduction in adverse cardiovascular outcomes [54].

Interventional Pharmacoepidemiology

Once the tools of pharmacoepidemiology make it possible to define patterns of suboptimal use such as poor drug choices, underuse, overuse, problematic dosing, and concomitant use of interacting drugs, such surveillance can be employed to identify problems amenable to interventions to improve utilization. Although epidemiology is traditionally seen as a merely observational discipline, it can also be used for what might be called "interventional epidemiology" – in this case, using the tools of pharmacoepidemiology to define baseline medication use, to direct the implementation of programs to improve such use, and then to employ the same rigorous ascertainment of practice patterns and clinical events to evaluate the effectiveness of those interventions.

One example of such interventional pharmacoepidemiology has been the development, testing, and widespread deployment of the form of educational outreach known as "academic detailing," discussed in greater detail in Chapter 19. This approach was designed to address observational data showing that prescribing patterns often appear to be shaped by the promotional efforts of drug manufacturers more strongly than by evidence-based guidelines. This is in large part because drug companies are much more effective in communicating their messages about what clinicians should prescribe than are academics. Much of industry's successful behavior change results from the activities of pharmaceutical sales representatives, known as "detailers," who go to the physician's office and engage in interactive conversations with the clinician that are specifically designed to change prescribing behavior. By contrast, most traditional continuing medical education offered by the academic world is far more passive: the physician is expected to come to a central location to attend a didactic presentation, usually with little interaction or feedback, and no clear-cut behavioral goal.

In the early 1980s, the academic detailing approach was developed, which used the engaging interactive outreach of the pharmaceutical industry, but put it in the service of transmitting messages based solely on evidence-based recommendations of optimal prescribing, developed by academic physicians [55]. Building on pharmacoepidemiologic assessment of overall prescribing patterns in a given area, the method was then tested in several population-based randomized trials in which it was shown to be effective in improving prescribing, as well as in reducing unnecessary medication expenditures [56–58].

The first academic detailing programs represent some of the earliest uses of population-based medication use datasets (in this case, from US Medicaid programs) to define medication use by large and well-defined populations of practitioners and patients. The availability of complete data on actual claims from the pharmacy datasets made possible a rigorous assessment of the interventions' efficacy as well as of their cost-effectiveness. Based on these initial observations, such programs have been subjected to over 60 subsequent randomized trials, and are now in widespread use globally.

As computerized drug and medical data have matured, their role has expanded to support large-scale, multicenter, pragmatic randomized trials of medications themselves [59]. In the Salford Lung Study, investigators randomized over 4000 typical patients with asthma to receive an inhaled combination of a beta-agonist and a corticosteroid or to usual care [60]. The trial was conducted in more than 70 general practice clinics in the UK, using an integrated electronic health record system that

enabled the investigators to collect study data during the course of the trial, with little additional interaction required between patients and trial staff.

Economic Assessment of Medication-Related Issues

Using population-based datasets that contain information on expenditures as well as utilization makes it possible to assess the economic impact of such prescribing issues as well (see Chapter 34). The previously mentioned study of patients treated for hypertension, for example, found that better adherence to the JNC guideline recommendations would not only have led to more evidence-based prescribing (and therefore better clinical outcomes), it would also have resulted in savings of $1.2 billion annually if the findings were projected nationally [12]. Similarly, the clopidogrel use study suggested that if aspirin had been substituted in patients who lacked an evidence-based or FDA-approved indication for use of the more costly drug, it would have saved $1.5 billion at a national level [13].

Another important application of pharmacoepidemiology to the economic assessment of medications builds on its capacity to model the effects of clinical trials well beyond their often brief duration [61]. For example, although statins usually must be taken for a lifetime, many randomized trials demonstrating their benefit have lasted for a much shorter time, often under two years. Epidemiologic methods make it possible to project the likely trajectories of simulated study subjects in both the experimental and control arms of a study. Based on differences observed during the trial itself, and some assumptions about their durability – assumptions which should be both transparent and conservative – it becomes possible to estimate the lifelong benefits, risks, and costs of use of such treatments [62].

The Academic Medical Center

The academic medical center represents a special case of inquiry for pharmacoepidemiology, and one where the field can make particularly useful contributions. These centers are the home base for many researchers in the field, and such settings are more likely than many routine practices to have available the electronic datasets that make such analyses possible. In recent years, the Institute of Medicine has been promoting the idea of a Learning Healthcare System in which the data generated within a medical center are analyzed and used to improve the delivery of care within the system. The science of pharmacoepidemiology is central to the collection, analysis, and interpretation of the data generated and used in this continuous feedback loop for several reasons, including its capacity to rigorously specify treatment exposures and outcomes, and its perspective that takes into account the concept of "population at risk" [63].

The application of population-based approaches can make it possible to subject problematic prescribing in an academic medical center to data-guided interventions, particularly if a computer-based order-entry system is being used (see Chapter 41) [64]. Until recently, this was possible only in advanced comprehensive healthcare organizations. However, in any institution in which prescriptions are written on a computerized order-entry system, prompts can be installed to propose more evidence-based medication use [65]. In addition, academic detailing programs or other interventions can then be deployed to address specific prescribing problems, and evaluated using the same order-entry data [57]. For academic medical centers that evolve in the coming years to become the hubs of comprehensive accountable care organizations, the availability of such data and investigator teams will make it possible to use these epidemiologic tools to study – and improve – the patterns of use and outcomes of medications across the entire inpatient–outpatient continuum of care.

Consortia of Academic Medical Center Programs for Pharmacoepidemiologic Research

As the field of pharmacoepidemiology matures, new collaborations are emerging to enhance the capacity of the healthcare delivery system and of academic centers to address important questions in medication use. Such collaborations can bring together large groups of patients for study, increasing the size of populations available for research, as well as their diversity and representativeness (see Chapter 25). Equally importantly, such consortia can bring together the expertise of several groups whose skills may be complementary in addressing the difficult methodologic issues inherent in observational studies of drug use and outcomes. The European Medicines Agency has created ENCePP, the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance [66]. The project has developed an inventory of European research centers and data sources in pharmacoepidemiology and pharmacovigilance, and provides a public index of such resources. ENCePP has also developed an electronic register of studies that provides a publicly accessible means of identifying all registered ongoing projects in pharmacoepidemiology and pharmacovigilance. In order to be registered and receive formal ENCePP approval, study investigators must agree to a Code of Conduct [67], which sets forth a set of principles for such studies concerning methodologic practices and transparency; they must also agree to adhere to a checklist of methodologic standards [68].

Examples in the US include the FDA's Sentinel, the Center for Disease Control and Prevention's (CDC) Vaccine Safety Datalink (VSD), and the Patient-Centered Outcome Research Institute's (PCORI) PCORnet. Sentinel is the FDA's national monitoring system that brings together a large number of electronic healthcare data providers and academic investigators to conduct postapproval safety surveillance of FDA-regulated medical products. The VSD, which is a collaborative project between the CDC and healthcare organizations that provide data and scientific expertise, is a precursor to Sentinel that focuses on vaccine safety surveillance, such as monitoring the safety of the seasonal influenza vaccine. A product of the healthcare reform program enacted in 2010, PCORI was designed to provide funding for comparative effectiveness research (CER), which was to include the study of medications, often by means of observational studies (see Chapter 26). PCORnet is PCORI's collaborative network of health systems, clinicians, researchers, patients, and data intended to foster patient-centered research across various health systems. However, those who expected PCORI to function as a CER resource that would fund trial or observational studies comparing relevant treatment options head to head have been surprised at what a small proportion of its activities have supported such work.

The Future

The continuing evolution of healthcare systems in both the industrialized and developing worlds will bring about a growing role for pharmacoepidemiology in multiple settings. Many new medications have novel efficacy but also daunting risks of toxicity, and often enormous costs. Healthcare systems all over the world face pressures to provide only those interventions that have the best efficacy and safety, but also at a price. To accomplish this will require relying on more than manufacturers' assessments of the utility, safety, or economic value of their own products, and more than clinicians' received wisdom or traditional prescribing habits. Nor will the interest of some insurers in promoting use of the most inexpensive medications necessarily

lead to optimal outcomes clinically, economically, or ethically. Pharmacoepidemiology (and its related discipline, pharmacoeconomics) can provide the tools for rigorous assessment of the good and harm that specific medications provide, and hold the promise of applying science to therapeutic decisions that are still too dominated by other forces.

References

1 Goldberg NH, Schneeweiss S, Kowal MK, Gagne JJ. Availability of comparative efficacy data at the time of drug approval in the United States. *JAMA* 2011; **305**(17): 1786–9.

2 Avorn J. *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. New York: Alfred A. Knopf, 2004.

3 Ross JS, Madigan D, Hill KP, Egilman DS, Wang Y, Krumholz HM. Pooled analysis of rofecoxib placebo-controlled clinical trial data: lessons for postmarket pharmaceutical safety surveillance. *Arch Intern Med* 2009; **169**(21): 1976–85.

4 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; **356**(24): 2457–71.

5 Williams SB, O'Connor E, Eder M, Whitlock E. Screening for child and adolescent depressions. In: *Primary Care Settings: A Systematic Evidence Review for the U.S. Preventive Services Task Force*. Rockville, MD: Agency for Healthcare Research and Quality, 2009.

6 Mostaghim SR, Gagne JJ, Kesselheim AS. Safety related label changes for new drugs after approval in the US through expedited regulatory pathways: retrospective cohort study. *BMJ* 2017; **358**: j3837.

7 Pinnow E, Amr S, Bentzen SM, *et al*. Postmarket safety outcomes for new molecular entity (NME) drugs approved by the Food and Drug Administration between 2002 and 2014. *Clin Pharmacol Ther* 2018; **104**(2): 390–400.

8 Kesselheim AS, Avorn J. New "21st Century Cures" legislation: speed and ease vs science. *JAMA* 2017; **317**(6): 581–2.

9 Fralick M, Kesselheim AS, Avorn J, Schneeweiss S. Use of health care databases to support supplemental indications of approved medications. *JAMA Intern Med* 2018; **178**(1): 55–63.

10 Setoguchi S, Glynn RJ, Avorn J, Levin R, Winkelmayer WC. Ten-year trends of cardiovascular drug use after myocardial infarction among community-dwelling persons > or = 65 years of age. *Am J Cardiol* 2007; **100**(7): 1061–7.

11 Solomon DH, Finkelstein JS, Katz JN, Mogun H, Avorn J. Underuse of osteoporosis medications in elderly patients with fractures. *Am J Med* 2003; **115**(5): 398–400.

12 Fischer MA, Avorn J. Economic implications of evidence-based prescribing for hypertension: can better care cost less? *JAMA* 2004; **291**: 1850–6.

13 Choudhry NK, Levin R, Avorn J. The economic consequences of non-evidence-based clopidogrel use. *Am Heart J* 2008; **155**(5): 904–9.

14 Cutrona SL, Choudhry NK, Stedman M, *et al*. Physician effectiveness in interventions to improve cardiovascular medication adherence: a systematic review. *J Gen Intern Med* 2010; **25**(10): 1090–6.

15 Setoguchi S, Choudhry NK, Levin R, Shrank WH, Winkelmayer WC. Temporal trends in adherence to cardiovascular medications in elderly patients after hospitalization for heart failure. *Clin Pharmacol Ther* 2010; **88**(4): 548–54.

16 Gurwitz JH, Glynn RJ, Monane M, *et al*. Treatment for glaucoma: adherence by the

elderly. *Am J Public Health* 1993; **83**(5): 711–16.

17 Monane M, Bohn RL, Gurwitz JH, Glynn RJ, Levin R, Avorn J. The effects of initial drug choice and comorbidity on antihypertensive therapy compliance: results from a population-based study in the elderly. *Am J Hypertens* 1997; **10**(7 Pt 1): 697–704.

18 Avorn J, Monette J, Lacour A, *et al*. Persistence of use of lipid-lowering medications: a cross-national study. *JAMA* 1998; **279**(18): 1458–62.

19 Jackevicius CA, Mamdani M, Tu JV. Adherence with statin therapy in elderly patients with and without acute coronary syndromes. *JAMA* 2002; **288**(4): 462–7.

20 Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. *Circulation* 2008; **117**(8): 1028–36.

21 Hakim A. E-prescribing and primary noncompliance: physician and patient experience. *Prof Case Manag* 2010; **15**(2): 62–7.

22 Fischer MA, Stedman MR, Lii J, *et al*. Primary medication non-adherence: analysis of 195,930 electronic prescriptions. *J Gen Intern Med* 2010; **25**(4): 284–90.

23 Raebel MA, Ellis JL, Carroll NM, *et al*. Characteristics of patients with primary non-adherence to medications for hypertension, diabetes, and lipid disorders. *J Gen Intern Med* 2012; **27**(1): 57–64.

24 Pottegård A, Christensen Rd, Houji A, *et al*. Primary non-adherence in general practice: a Danish register study. *Eur J Clin Pharmacol* 2014; **70**(6): 757–63.

25 Thengilsdóttir G, Pottegård A, Linnet K, Halldórsson M, Almarsdóttir AB, Gardarsdóttir H. Do patients initiate therapy? Primary non-adherence to statins and antidepressants in Iceland. *Int J Clin Pract* 2015; **69**(5): 597–603.

26 da Costa FA, Pedro AR, Teixeira I, Bragança F, da Silva JA, Cabrita J. Primary non-adherence in Portugal: findings and implications. *Int J Clin Pharm* 2015; **37**(4): 626–35.

27 Chan DC, Shrank WH, Cutler D, *et al*. Patient, physician, and payment predictors of statin adherence. *Med Care* 2010; **48**(3): 196–202.

28 Shrank WH, Hoang T, Ettner SL, *et al*. The implications of choice: prescribing generic or preferred pharmaceuticals improves medication adherence for chronic conditions. *Arch Intern Med* 2006; **166**(3): 332–7.

29 Gagne JJ, Choudhry NK, Kesselheim AS, *et al*. Comparative effectiveness of generic and brand-name statins on patient outcomes: a cohort study. *Ann Intern Med* 2014; **161**(6): 400–7.

30 Curtis JR, Xi J, Westfall AO, *et al*. Improving the prediction of medication compliance: the example of bisphosphonates for osteoporosis. *Med Care* 2009; **47**(3): 334–41.

31 Solomon DH, Brookhart MA, Tsao P, *et al*. Predictors of very low adherence with medications for osteoporosis: towards development of a clinical prediction rule. *Osteoporos Int* 2011; **22**(6): 1737–43.

32 Hill TD, Graham LM, Divgi V. Racial disparities in pediatric asthma: a review of the literature. *Curr Allergy Asthma Rep* 2011; **11**(1): 85–90.

33 Pitt B, Zannad F, Remme WJ, *et al*. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med* 1999; **341**(10): 709–17.

34 Juurlink DN, Mamdani MM, Lee DS, *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; **351**(6): 543–51.

35 Choudhry NK, Zagorski B, Avorn J, *et al*. Comparison of the impact of the Atrial Fibrillation Follow-Up Investigation of Rhythm Management trial on prescribing patterns: a time-series analysis. *Ann Pharmacother* 2008; **42**(11): 1563–72.

36 Gurwitz JH, Monette J, Rochon PA, Eckler MA, Avorn J. Atrial fibrillation and stroke prevention with warfarin in the long-term care setting. *Arch Intern Med* 1997; **157**(9): 978–84.

37 Choudhry NK, Anderson GM, Laupacis A, Ross-Degnan D, Normand SL, Soumerai SB. Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis. *BMJ* 2006; **332**(7534): 141–5.

38 Monette J, Gurwitz JH, Rochon PA, Avorn J. Physician attitudes concerning warfarin for stroke prevention in atrial fibrillation: results of a survey of long-term care practitioners. *J Am Geriatr Soc* 1997; **45**(9): 1060–5.

39 Parikh S, Mogun H, Avorn J, Solomon DH. Osteoporosis medication use in nursing home patients with fractures in 1 US state. *Arch Intern Med* 2008; **168**(10): 1111–15.

40 Jackevicius CA, Tu JV, Ross JS, Ko DT, Krumholz HM. Use of ezetimibe in the United States and Canada. *N Engl J Med* 2008; **358**(17): 1819–28.

41 Shah ND, Montori VM, Krumholz HM, Tu K, Alexander GC, Jackevicius CA. Responding to an FDA warning—geographic variation in the use of rosiglitazone. *N Engl J Med* 2010; **363**(22): 2081–4.

42 Najafzadeh M, Schneeweiss SS, Choudhry NK, Wang SV, Gagne JJ. Simulation for predicting effectiveness and safety of new cardiovascular drugs in routine care populations. *Clin Pharmacol Ther* 2018; **104**(5): 1008–15.

43 Schneeweiss S, Maclure M, Carleton B, Glynn RJ, Avorn J. Clinical and economic consequences of a reimbursement restriction of nebulised respiratory therapy in adults: direct comparison of randomised and observational evaluations. *BMJ* 2004; **328**(7439): 560.

44 Choudhry NK, Fischer MA, Avorn J, *et al*. At Pitney Bowes, value-based insurance design cut copayments and increased drug adherence. *Health Aff (Millwood)* 2010; **29**(11): 1995–2001.

45 Choudhry NK, Avorn J, Glynn RJ, *et al*.; Post-Myocardial Infarction Free Rx Event and Economic Evaluation (MI FREEE) Trial. Full coverage for preventive medications after myocardial infarction. *N Engl J Med* 2011; **365**(22): 2088–97.

46 Matlin OS, Kymes SM, Averbukh A, *et al*. Community pharmacy automatic refill program improves adherence to maintenance therapy and reduces wasted medication. *Am J Manag Care* 2015; **21**(11): 785–91.

47 Fischer MA, Choudhry NK, Bykov K, *et al*. Pharmacy-based interventions to reduce primary medication nonadherence to cardiovascular medications. *Med Care* 2014; **52**(12): 1050–4.

48 Krumme AA, Glynn RJ, Schneeweiss S, *et al*. Medication synchronization programs improve adherence to cardiovascular medications and health care use. *Health Aff (Millwood)* 2018; **37**(1): 125–33.

49 Fischer MA, Schneeweiss S, Avorn J, Solomon DH. Medicaid prior-authorization programs and the use of cyclooxygenase-2 inhibitors. *N Engl J Med* 2004; **351**(21): 2187–94.

50 Polinski JM, Wang PS, Fischer MA. Medicaid's prior authorization program and access to atypical antipsychotic medications. *Health Aff (Millwood)* 2007; **26**(3): 750–60.

51 Fischer MA, Choudhry NK, Winkelmayer WC. Impact of Medicaid prior authorization on angiotensin-receptor blockers: can policy promote rational prescribing? *Health Aff (Millwood)* 2007; **26**(3): 800–7.

52 Fischer MA, Servi AD, Polinski JM, Wang PS. Restrictions on antidepressant medications for children: a review of Medicaid policy. *Psychiatr Serv* 2007; **58**(1): 135–8.

53 Lu CY, Soumerai SB, Ross-Degnan D, Zhang F, Adams AS. Unintended impacts of a Medicaid prior authorization policy on access to medications for bipolar illness. *Med Care* 2010; **48**(1): 4–9.

54 Jackevicius CA, Tu JV, Demers V, *et al*. Cardiovascular outcomes after a change in prescription policy for clopidogrel. *N Engl J Med* 2008; **359**(17): 1802–10.

55 Avorn J, Soumerai SB. Improving drug-therapy decisions through educational outreach: a randomized controlled trial of academically based "detailing." *N Engl J Med* 1983; **308**(24): 1457–63.

56 Avorn J, Soumerai SB, Everitt DE, *et al*. A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. *N Engl J Med* 1992; **327**(3): 168–73.

57 Solomon DH, Van Houten L, Glynn RJ, *et al*. Academic detailing to improve use of broad-spectrum antibiotics at an academic medical center. *Arch Intern Med* 2001; **161**(15): 1897–902.

58 Soumerai SB, Avorn J. Economic and policy analysis of university-based drug "detailing." *Med Care* 1986; **24**(4): 313–31.

59 Choudhry NK. Randomized, controlled trials in health insurance systems. *N Engl J Med* 2017; **377**(10): 957–64.

60 Vestbo J, Leather D, Diar Bakerly N, *et al*.; Salford Lung Study Investigators. Effectiveness of fluticasone furoate-vilanterol for COPD in clinical practice. *N Engl J Med* 2016; **375**(13): 1253–60.

61 Ganz DA, Kuntz KM, Jacobson GA, Avorn J. Cost-effectiveness of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor therapy in older patients with myocardial infarction. *Ann Intern Med* 2000; **132**(10): 780–7.

62 Choudhry NK, Patrick AR, Glynn RJ, Avorn J. The cost-effectiveness of C-reactive protein testing and rosuvastatin treatment for patients with normal cholesterol levels. *J Am Coll Cardiol* 2011; **57**(7): 784–91.

63 Schneeweiss S. Learning from big health care data. *N Engl J Med* 2014; **370**(23): 2161–3.

64 Schedlbauer A, Prasad V, Mulvaney C, *et al*. What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? *J Am Med Inform Assoc* 2009; **16**(4): 531–8.

65 Jha AK, Laguette J, Seger A, Bates DW. Can surveillance systems identify and avert adverse drug events? A prospective evaluation of a commercial application. *J Am Med Inform Assoc* 2008; **15**(5): 647–53.

66 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *About ENCePP*. http://www.encepp.eu/structure/index.shtml (accessed March 2018).

67 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *Code of Conduct*. http://www.encepp.eu/code_of_conduct/index.shtml (accessed March 2018).

68 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *Standards and Guidance*. http://www.encepp.eu/standards_and_guidances/index.shtml (accessed March 2018).

Further Reading

Avorn J. *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. New York: Knopf, 2005.

Avorn J. Teaching clinicians about drugs—50 years later, whose job is it? *N Engl J Med* 2011; **364**: 1185–7.

Avorn J, Soumerai SB. Improving drug-therapy decisions through educational outreach. A randomized controlled trial of academically based "detailing." *N Engl J Med* 1983; **308**: 1457–63.

Choudhry NK, Avorn J, Glynn RJ, *et al*.; Post-Myocardial Infarction Free Rx Event and

Economic Evaluation (MI FREEE) Trial. Full coverage for preventive medications after myocardial infarction. *N Engl J Med* 2011; **365**: 2088–97.

Clancy C, Collins FS. Patient-Centered Outcomes Research Institute: the intersection of science and health care. *Sci Transl Med* 2010; **2**: 37.

Cutrona SL, Choudhry NK, Stedman M, *et al*. Physician effectiveness in interventions to improve cardiovascular medication adherence: a systematic review. *J Gen Intern Med* 2010; **25**: 1090–6.

Fischer MA, Avorn J. Economic implications of evidence-based prescribing for hypertension: can better care cost less? *JAMA* 2004; **291**: 1850–6.

Fischer MA, Choudhry NK, Winkelmayer WC. Impact of Medicaid prior authorization on angiotensin-receptor blockers: can policy promote rational prescribing? *Health Aff (Millwood)* 2007; **26**: 800–7.

Fischer MA, Morris CA, Winkelmayer WC, Avorn J. Nononcologic use of human recombinant erythropoietin therapy in hospitalized patients. *Arch Intern Med* 2007; **167**: 840–6.

Fischer MA, Stedman MR, Lii J, *et al*. Primary medication non-adherence: analysis of 195,930 electronic prescriptions. *J Gen Intern Med* 2010; **25**: 284–90.

Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. *Circulation* 2008; **117**: 1028–36.

Jackevicius CA, Tu JV, Demers V, *et al*. Cardiovascular outcomes after a change in prescription policy for clopidogrel. *N Engl J Med* 2008; **359**: 1802–10.

Juurlink DN, Mamdani MM, Lee DS, *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; **351**: 543–51.

O'Brien MA, Rogers S, Jamtvedt G, *et al*. Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007; **4**: CD000409.

Shah ND, Montori VM, Krumholz HM, Tu K, Alexander GC, Jackevicius CA. Responding to an FDA warning—geographic variation in the use of rosiglitazone. *N Engl J Med* 2010; **363**: 2081–4.

Shrank WH, Hoang T, Ettner SL, *et al*. The implications of choice: prescribing generic or preferred pharmaceuticals improves medication adherence for chronic conditions. *Arch Intern Med* 2006; **166**: 332–7.

Solomon DH, Finkelstein JS, Katz JN, Mogun H, Avorn J. Underuse of osteoporosis medications in elderly patients with fractures. *Am J Med* 2003; **115**: 398–400.

Solomon DH, Van Houten L, Glynn RJ, *et al*. Academic detailing to improve use of broad-spectrum antibiotics at an academic medical center. *Arch Intern Med* 2001; **161**: 1897–902.

7

The Role of Pharmacoepidemiology in Industry

Nicolle M. Gatto[1,2,3], Rachel E. Sobel[1], Jamie Geier[1], Jingping Mo[1], Andrew Bate[4,5], and Robert F. Reynolds[3,6]

[1] *Department of Epidemiology, Worldwide Research and Development, Pfizer Inc., New York, NY, USA*
[2] *Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA*
[3] *Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA*
[4] *Pfizer Ltd, Walton-on-the-Hill, Surrey, UK*
[5] *Department of Medicine, NYU School of Medicine, New York, NY, USA*
[6] *Department of Epidemiology, Global Medical Value, Evidence and Outcomes, GlaxoSmithKline, Upper Providence, PA, USA*

Epidemiology is recognized as a key component of risk management and safety assessment activities during pre- and postapproval drug development. In addition to risk management, epidemiology contributes to several other important functions within a biopharmaceutical company, including portfolio development, studies of beneficial drug effects, the commercialization of drugs, and benefit–risk assessments. The use of epidemiology to support the appropriate marketing of drugs, including studies of beneficial drug effects, health economics, quality-of-life measures, and benefit–risk assessment, are discussed elsewhere in this book (see Chapters 26, 33, 34, 35, and 42). The most consistent contribution of epidemiology in the biopharmaceutical industry is arguably drug safety evaluation, including the contextualization and refinement of safety signals, and examination of specific research hypotheses. To meet these aims, epidemiologists design and implement background epidemiology studies among indicated populations, risk management interventions and evaluations, and postapproval safety studies. Additionally, epidemiologists contribute content, expertise, and strategy to regulatory documents such as global risk management plans (RMP), pediatric investigation plans (PIP), and orphan drug applications, and are key contributors in interactions with regulatory authorities.

This chapter discusses the specific application of pharmacoepidemiology to safety assessment throughout the development life cycle, from the perspective of epidemiologists working within the biopharmaceutical industry. Throughout this chapter we will refer as an example to the epidemiology strategy implemented to support the development, approval, and postapproval activities for Xeljanz® (tofacitinib), a Janus kinase (JAK) inhibitor for treatment of rheumatoid arthritis (RA).

The views expressed are those of the authors, which are not necessarily those of Pfizer Inc.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Regulatory and Industry Focus on Risk Management and Epidemiology

Biopharmaceutical risk management (see also Chapter 24) is fundamentally concerned with preserving an appropriate benefit–risk balance among patients using a medicine, vaccine, or device. There are many tools by which this goal can be achieved, but *risk assessment* and *risk mitigation* are the two primary components of risk management. Epidemiologists play a vital role in the quantification and interpretation of risk. Preapproval, they contextualize risks emerging from clinical studies by understanding the background rates of disease occurrence in the indicated population. Postapproval, they assess the safety of drugs as used in actual clinical practice. Epidemiologists' training in observational research, data analysis and interpretation, and survey and program design also contributes to effective risk mitigation program planning and assessment.

The Evolution of Biopharmaceutical Risk Management

The guidance and regulations around risk management have evolved since the 1990s. Public pressure to speed up drug approvals for HIV and cancer drugs led to the Prescription Drug User Fee Act (PDUFA) in the US. Ten years later, concern that speed might come at the expense of fully evaluating safety led to the inclusion of a risk management framework for safety assessment in PDUFA III in 2002. For the first time, dedicated funding was provided to the Food and Drug Administration (FDA) for risk management resources. In response to this regulation, the FDA issued three guidance documents in 2005: (i) Pre-Marketing Risk Assessment, (ii) Pharmacovigilance and Pharmacoepidemiology, and (iii) Risk Minimization Action Plans (RiskMAPs).

After a number of widely used drugs were withdrawn in 2004 and 2005 for safety reasons, the public questioned the effectiveness of the FDA's methods of assessing and approving drugs. The Institute of Medicine (IOM) was tasked with evaluating the US drug safety system and making recommendations for improvements to risk assessment, safety surveillance, and the safer use of drugs. The IOM committee made numerous recommendations, several of which pertained to epidemiologists, including that the FDA receive additional funding and staff; improve communications on drug safety, incorporating a larger role for drug safety staff; and, most importantly, be given additional authority and enforcement tools [1].

As a result of the IOM report and other stakeholder research and advocacy, Congress passed the Food and Drug Administration Amendment Act (FDAAA) in 2007, which further strengthened the FDA's oversight of risk management activities. With FDAAA, the FDA was granted the ability to mandate postapproval studies (postmarketing requirements, or PMR) and risk evaluation and mitigation strategies (REMS; see the later section for further information) by imposing substantial fines for noncompliance or denial/revocation of drug approval. FDAAA also allowed for voluntary postmarketing commitments (PMC); that is, studies that may not necessarily be required, but could provide important public health information. Observational studies could be either PMRs or PMCs, and are further described in the FDA Guidance for Industry Postmarketing Studies and Clinical Trials [2].

Europe passed similar legislation in 2005, the Rules Governing Medicinal Products in the European Union, Volume 9A, which provide guidelines on pharmacovigilance and risk management between companies and the European Medicines Agency (EMA) [3]. European Union (EU) law requires companies to submit a formal RMP with each marketing authorization application (MAA). Following a review of the

European system of safety monitoring as well as extensive public consultation, a Directive and Regulation (also called the new EU pharmacovigilance legislation) were adopted by the European Parliament and Council of Ministers in December 2010, which became effective in July 2012, bringing about significant changes in the safety monitoring of medicines across the EU. The new EU pharmacovigilance legislation introduced a pharmacovigilance system master file (PSMF), required RMPs for all new products, enhanced postauthorization measures with legally binding postauthorization safety studies (PASS), including evaluation of the effectiveness of additional risk minimization measures (aRMMs), and postauthorization efficacy studies (PAES). The new EU pharmacovigilance legislation also introduced clarity in the oversight by the authorities for noninterventional studies: the national competent authority is responsible for nationally authorized products; EMA and its Pharmacovigilance and Risk Assessment Committee (PRAC) has oversight responsibility when more than one member state is involved. To facilitate the performance of pharmacovigilance in accordance with the new EU legislation, the EMA developed good pharmacovigilance practices (GVP) modules. The modules that are most relevant to epidemiologists are Module VIII – Post-authorization safety studies [4] and Module XVI – Risk minimization measures: selection of tools and effectiveness [5].

Besides the US and EU, regulations on risk management planning, including postapproval safety studies, are evolving in other parts of the world, such as Asia and Latin America. In Japan, postmarketing surveillance (PMS) is required for newly approved medicine and must be conducted in according with good postmarketing study practice (GPSP), a set of standards unique to Japan. The GPSP ordinance mandates PMS studies, commonly known as drug use results surveys (DURS), and defines the approach for the conduct of DURS. There is a little flexibility in design and format, and protocol finalization and approval are usually streamlined processes. Japan's Pharmaceuticals and Medical Devices Agency (PMDA) has been working to strengthen its drug safety assessment framework. The Medical Information for Risk Assessment Initiative (MIHARI) project was initiated in 2009 with the aim of utilizing large-scale electronic health information databases as novel information sources for pharmacoepidemiologic drug safety assessments in Japan. After conducting extensive pilot studies, the framework was implemented in practical applications in 2014, and is expected to play an important role in Japan's pharmacovigilance and risk management in the future.

In China, policies for postapproval safety studies, known as *intensive monitoring,* are still evolving, and the available guidance and overall approach are not as comprehensive as in the US, EU, or Japan. However, the basis for intensive monitoring has evolved over the past decade, and provisions for the ideas of postmarketing reevaluation and reregistration are delineated in China's Food and Drug Administration regulations.

In Mexico, the Federal Commission for Protection against Health Risks (COFEPRIS) is the regulatory authority for pharmaceuticals. The National Center of Pharmacovigilance within COFEPRIS is responsible for the oversight of all pharmacovigilance activities, in addition to setting local policies in line with national and international pharmacovigilance guidelines. The main standard guideline governing pharmacovigilance, including PMS studies, in Mexico is the Installation and Operation of Pharmacovigilance (2013).

Epidemiology has become increasingly important to risk management over the last three decades with the evolution of pharmacovigilance regulation globally, which has further solidified epidemiology's role in informing the benefit–risk assessment of medicines throughout the development lifecycle.

Epidemiology in Drug Safety Evaluation

Background

The safety profile of any drug reflects an evolving body of knowledge, extending from preclinical investigations through to the postapproval life cycle of the product. Drug manufacturers traditionally relied on two major sources for information on the safety of drugs: the clinical trials supporting the new drug application (NDA) and, once the drug was marketed, spontaneous reports received throughout the world (see Chapter 10). Clinical trials and spontaneous reports are useful and have a unique place in assessing drug safety (e.g., signal detection). However, both sources have limitations that can be addressed, in part, by the proper use of observational epidemiology. Epidemiologic studies complement these two sources of data to refine the safety signals they generate and to provide a more comprehensive and pragmatic picture of the safety profile of a drug as it is used in clinical practice.

Pharmacoepidemiology Study Designs

Pharmacoepidemiologic analyses can be descriptive or analytic in nature; may involve existing data, primary data collection, or a hybrid of secondary and primary data; and may be used to generate, refine, or examine hypotheses. Industry epidemiologists compile drug safety information from published epidemiologic literature, pooled clinical trials, trial extensions, electronic health records (e.g., insurance claims data or electronic medical records), existing registries, and *de novo* observational studies. Commonly used study designs include the cohort study, case–control study, and cross-sectional study (see also Chapter 3).

In addition to typical epidemiologic designs, depending on the specific safety research hypothesis, epidemiologists design and implement active surveillance studies, pragmatic trials (including the most naturalistic version, the large simple trial), and self-controlled designs such as the case-crossover study and self-controlled case series. Active surveillance studies can be defined as descriptive studies intended to solicit information on adverse events among a specified population, such that the numerator and denominator are as complete as possible, potentially allowing calculation of incidence. An example describing the active surveillance program established for juvenile idiopathic arthritis can be found later in this chapter in the section on pediatrics.

Purely observational epidemiologic studies may not always be the most appropriate method of evaluating safety signals or comparing the safety profile of different medications, especially when there are concerns of confounding by indication. Confounding by indication occurs when the risk of an adverse event is related to the indication for medication use, such that in the absence of the medication, those actually exposed are at higher or lower risk of the adverse event than those unexposed. As with any other form of confounding, one can, in theory, control for its effects if the severity of the underlying illness (i.e., any conditions specified as labeled indications or contraindications, or included in the precautions or warnings) can be validly measured (see Chapter 43). Confounding by indication is more of an issue when a particular property of the drug is very likely to affect the type of patient it is used by or prescribed to. In these cases, studies using randomization to treatment may be necessary. A pragmatic clinical trial (PCT) is a randomized clinical trial with one or more pragmatic elements, and a large simple trial (LST) is a type of PCT that combines randomization to treatment with observational follow-up of patients. The LST design allows for a theoretical balance of known and unknown confounding factors, while maintaining more real-world safety assessment than typical

clinical trials. By maintaining simplicity in study procedures, including the study's inclusion/exclusion criteria, patients' use of concomitant medications, and the frequency of patient monitoring, LSTs approximate real-life practice. Further, the large study size provides the power needed to evaluate small absolute and relative risks. The characteristics of LSTs are further described in Chapter 32.

Self-controlled designs, in which each case serves as its own control, were developed to assess effects of intermittent exposures on abrupt-onset events or diseases (see also Chapter 43). The case-crossover study is one form of self-controlled design, analogous to a traditional matched case–control design. The risk window is defined as a time period just before the outcome occurred and the control window(s) is (are) defined as other (nonoverlapping) time periods before the outcome occurred. For example, the case-crossover design was used to evaluate whether PDE5 inhibitors (i.e., sildenafil, vardenafil, and tadalafil), as a class, triggered the onset of acute nonarteritic anterior ischemic optic neuropathy (NAION) [6,7]. NAION is a rare condition, necessitating a design that identifies patients based on their disease status. However, identifying appropriate controls was deemed very challenging with a standard case–control study design. Furthermore, the case-crossover study was ideally suited to this research question, since PDE5 inhibitors are taken on an as-needed basis, which constitutes an intermittent exposure; acute NAION is characterized by sudden onset and is experienced by the patient as an abrupt visual change, often first detected upon awakening; and each case subject is effectively matched to him- or herself, such that the potential effects of confounders that do not vary over the study period, such as age, diabetes, and hypertension, are effectively held constant.

In the self-controlled case series design (SCCS), originally developed to study vaccine safety (see Chapter 20), the risk and control windows are defined within some observation period of interest. The risk window is defined as a time period after the exposure of interest occurs (among those exposed) and the control windows are all other time periods before and after the risk window (among those exposed), or all time periods (for those unexposed) within the observation period. The SCCS was used to evaluate the global risk of Guillain–Barré syndrome (GBS) following vaccination against Influenza A (H1N1) [8]. In this case, the SCCS design was chosen because it does not require the infrastructure needed to establish complete population denominators. In addition, the study design requirements were met, for instance H1N1 vaccine was an intermittent exposure, GBS is rare and has an abrupt onset, and the likelihood of H1N1 vaccine exposure was not expected to be impacted by the development of GBS [9].

Data Sources for Pharmacoepidemiology Studies

In order to respond rapidly and responsibly to safety issues, high-quality and valid data resources must be available. As a result of this need, the development and use of record linkage and automated databases, including hospital databases, have grown considerably over the past several decades (see also Chapters 11–14). These databases offer several advantages over primary data collection epidemiology studies or randomized trials (see Chapters 15, 16, and 32). First, automated databases are usually large in size, ranging from hundreds of thousands to millions of patients, often with many years of "observation." A second advantage is speed: since information on study subjects is already computerized, the data can be accessed quickly rather than waiting years for the results of studies in which patients are identified and followed over time. The third advantage is cost relative to primary data collection studies. Primary data collection observational studies and randomized trials can cost tens to hundreds of

millions of dollars, compared to hundreds of thousands of dollars for database studies.

Considerable progress has been made in the development of new and existing research databases containing information on drug usage and health-related outcomes. This progress is advantageous, as a variety of data sources are necessary for research in pharmacoepidemiology. The limitations of many automated datasets are well established and need to be considered before conducting a study on a newly marketed medication. Each data source will have its own strengths and limitations, which are usually related to important factors: the reasons for collecting the data (e.g., research, monitoring clinical practice, or reimbursement); the type of data collected; the coding systems used; the resources devoted to evaluating and monitoring the research quality of the data; and national or regional variations in medical practice. Many data collection systems were designed for administrative, rather than research, purposes. As a result, information needed to assess a specific safety issue may be unavailable and the quality of medical information may be inadequate. Validation of outcomes based on diagnostic or procedural codes used for reimbursement purposes (or algorithms based on these codes) against at least a subset of medical records is often desirable, as the usefulness of this type of research to answer an important safety question may be limited if the data are not properly validated. For some databases, medical record review may not be feasible due to laws and/or policies regarding patient confidentiality or anonymity. Continuing studies of the research validity of these databases are crucial, and should be pursued when feasible [10–14]. Further information on specific data sources can be found in Chapters 11–17.

Another common research limitation of automated data sources is that sufficient numbers of users may not yet be recorded, or the medication may not be marketed in the country where the database is located. Some data resources suffer from a considerable "lag time" between data entry and availability for research purposes. Further, even though many health maintenance organizations and national patient registries have a very large number of unique persons (often millions), these numbers may be inadequate to study potential drug risks of extremely rare outcomes. Finally, results from these sources are often limited in their generalizability.

Epidemiologic studies with primary data collection are considered when it is not feasible to address safety issues using existing databases. The NAION case-crossover studies described earlier and the ziprasidone LST (described in Chapter 32) are examples of epidemiology studies that involved primary data collection. These types of studies take a relatively long time for data collection and are more costly. In some instances, it is possible to conduct a hybrid primary–secondary data collection study by identifying patients or physicians in automated databases, and supplementing the existing data source with information collected directly from patients or physicians through telephone interviews or via electronic communications [15,16].

To address any specific product safety concerns, it is important to consider the validity and feasibility of all potential study design and data source options to enable selection of the most appropriate, and to implement epidemiologic studies in accordance with relevant guidelines such as the Guidelines for Good Pharmacoepidemiology Practices (GPP) [17], the EMA's European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology [18], and the FDA Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting of Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets [19].

Contributions of Preapproval Epidemiology

Before evaluation of a potential medicine can begin, extensive preclinical research is con-

ducted, involving lengthy *in vitro* and *in vivo* testing. Preclinical safety studies evaluate and identify potential toxic effects of the drug, which include assessing whether a medicine is carcinogenic, mutagenic, or teratogenic. Although the information generated from preclinical studies provides guidance on the selection of a safe starting dose for the first administration-to-human study, the limited predictability of animal studies to the toxicity of drugs in human is well recognized. However, these studies can provide important information about hypothetical drug risks.

Randomized clinical trials (RCTs) provide abundant high-quality data about identified and hypothetical risks, but have limitations. Preapproval RCTs typically involve highly selected subjects, followed for a short period of time, and in the aggregate include at most a few thousand patients. These studies are generally sufficiently large to provide evidence of a beneficial clinical effect, exclude large increases in risk of common adverse events, and identify the most common and acutely occurring adverse events. However, they are rarely large enough to detect small differences in the risk of common adverse events or to reliably estimate the risk of rare events. Using the "rule of three," where the sample size needed is roughly three times the reciprocal of the frequency of the event, at least 300 patients would be required in a trial in order to observe at least one adverse event that occurs at a rate of 1/100. Likewise, a sample of 3000 is needed to observe at least one adverse event with 95% probability if the frequency of the event is 1/1000. (See Chapter 4 for more discussion of the sample sizes needed for studies.) Increasingly, preapproval studies – particularly in rare diseases or where long-term placebo treatments are unethical – include unbalanced randomization or treatment arms with a short duration of placebo or active comparator, or use noncontemporaneous controls. While clinical trials are not intended or designed to address all potential safety issues related to a particular drug [20], like preclinical studies, they often give rise to signals that cannot be adequately addressed from trial data alone.

Preapproval epidemiology complements safety data from preclinical and clinical studies and provides a context for signals arising from clinical trials. Comprehensive reviews of the epidemiologic literature are complemented by epidemiologic studies to establish among patients expected to use the new medication (i.e., indicated populations) the background epidemiology (e.g., incidence, prevalence, mortality) of the indication; the expected prevalence/incidence of risk factors, co-morbidities, and complications; patterns of healthcare utilization and prescribing of currently approved treatments; and background rates of mortality and serious nonfatal events.

Epidemiologic studies conducted before or during the clinical development program are often critical to place the incidence of adverse events observed in clinical trials in perspective. Data are often lacking on the expected rates of events in the population likely to be treated. For example, studies examining the risk factors for and rates of sudden unexplained death among people with epilepsy were able to provide reassurance that the rates observed in a clinical development program were within the expected range for individuals with comparably severe disease [21–23]. Epidemiologists use information from the published literature, descriptive epidemiologic studies, and standing cohorts (i.e., open cohorts of indicated populations which are updated over time and queried for incidence of safety events and other data as needed) to support regulatory filings and to complete epidemiology sections of key regulatory documents (e.g., risk management and pediatric investigation plans, orphan drug applications). These background epidemiology data can also be a key component for internal decision making, such as trial design, data monitoring committee decisions to stop/continue trials, decisions to move/not move to the next phase

of development, risk management decisions, and risk mitigation planning.

During development, in addition to summarizing the existing relevant literature and designing and executing background epidemiology studies, industry epidemiologists are often involved in safety signal evaluation, observational analyses of RCT data (e.g., as treated or observed versus expected analyses), and designing postapproval epidemiology studies and risk minimization planning. Planning for successful postapproval epidemiology studies often begins well before approval. During the periapproval phase, epidemiologists may conduct feasibility assessments for planned postapproval studies, start key operational aspects of postapproval studies (e.g., identifying key external partners such as contract research organizations and scientific steering committee members for the design and conduct of the study), and contribute to regulatory submissions, responses, and negotiations (e.g., responding to regulatory inquiries related to epidemiology, participate in regulatory meetings).

There are several other areas where epidemiologists are increasingly providing their expertise to support preapproval development. In the context of risk minimization planning, the epidemiologist may conduct research to test the comprehension and utility of educational materials, evaluate the proposed risk minimization tools and processes to assess their burden on the healthcare system and patients, pilot and/or user test assessment materials such as surveys, and generally contribute to the design and implementation of these programs. Furthermore, many regulatory agencies are utilizing various benefit–risk assessment frameworks in their reviews. Epidemiologists can provide inputs or lead both quantitative and qualitative benefit–risk assessments such as multicriteria decision analysis (MCDA) [24], stochastic multicriteria acceptability analysis (SMAA) [25], and the PhRMA Benefit-Risk Action Team (BRAT) framework [26], among others (see Chapter 35). Lastly, several accelerated/conditional approval pathways and regulations exist in the EU, and are anticipated for the US and other regions, which have requirements for real-world data and evidence (RWD/RWE) to complement the incomplete or uncertain data from abbreviated development programs in areas of high unmet need. Epidemiologists' expertise in regulatory-quality RWD/RWE generation is often critical to the success of these accelerated options.

Example: Tofacitinib Preapproval Epidemiology Strategy

The epidemiologic strategy for tofacitinib incorporated several distinct but complementary efforts for risk characterization, including literature reviews, meta-analyses, and a standing cohort within a US-based registry of patients with RA (see Figure 7.1). While all RCTs in the tofacitinib RA development program included at least one control group (placebo or active), the size of the control groups and duration of treatment did not permit precise comparative assessments for adverse events with low frequency or long latency. Long-term extension (LTE) studies provided greater exposure in patients taking tofacitinib. However, the lack of a control group within the LTE studies precluded direct comparative risk assessment; interpretation, therefore, was difficult, in that no data were collected that could provide evidence of the expected rates in a concurrent and directly comparable patient population. The use of indirect comparative methods via external patient cohorts provided such context, while taking into account key potential differences in the populations compared, whose baseline demographic characteristics, disease course, and treatment history varied. Data from multiple sources (i.e., observational studies, RCTs with other agents, and cohorts of patients stratified by RCT inclusion/exclusion criteria) were used to provide these indirect comparisons, drawing from the strengths of each data source while balancing their weaknesses. The output of the analyses was used to assess the rates of

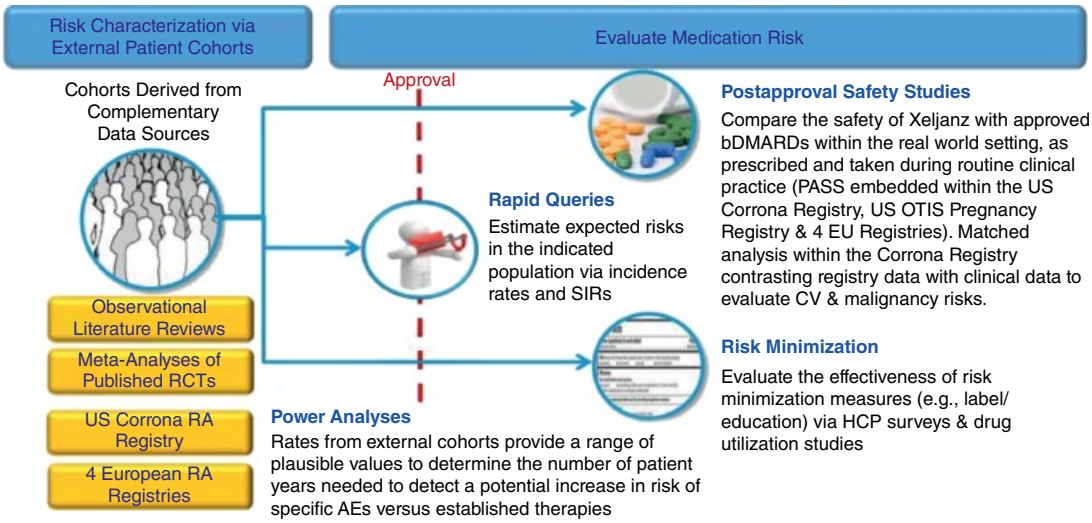


Figure 7.1 Harnessing the power of real-world evidence: Pharmacoepidemiology strategy for Xeljanz® (tofacitinib). AE, adverse event; CV, cardiovascular; HCP, healthcare professional; RA, rheumatoid arthritis; RCT, randomized controlled trial; SIR, standardized incidence rate.

identified and potential risks of interest in the tofacitinib clinical program compared with those from cohorts of RA patients treated with biologic disease-modifying antirheumatic drugs (bDMARDs). These efforts were articulated within regulatory documents, including the NDA for the FDA and the briefing document (i.e., summary of clinical safety [SCS], clinical overview [CO], etc.) for the EMA, and the data were also presented at the FDA Advisory Committee Meeting for tofacitinib (May 2012).

Due to the timing of approvals and regulatory interactions in the EU, these descriptive analyses were expanded to include four EU-based database studies. The strategy and accompanying power analysis were presented in several meetings with European national regulatory authorities and key EMA decision-makers. After deliberations, the EMA agreed to the inclusion of an integrated summary of the risk characterization work and key findings within the submission dossier (with analysis-specific component reports as appendices). The collective body of evidence (including interim data from the US-based registry study) provided substantial additional context to rates of selected adverse events observed in the tofacitinib clinical trial program, and therefore addressed the uncertainties related to the potential risks previously expressed by the EMA. Tofacitinib was approved in the EU in March 2017.

Contributions of Postapproval Epidemiology

The need for a postapproval epidemiology study can be known and devised preapproval or can arise once a new drug is marketed. Postapproval signals may come from clinical trial extension data (e.g., LTE studies), spontaneous reports, published case series, or signal detection of electronic healthcare data. Postapproval, epidemiologists execute postapproval commitments (e.g., epidemiology studies, active surveillance studies, other registries, REMS/aRMM evaluations, PIP observational studies, etc.); conduct studies evaluating the effectiveness of risk mitigation activities; perform signal detection in

existing cohorts (e.g., via claims or electronic patient record data); and design and implement new studies as additional signals arise (e.g., from spontaneous reports, signal detection, or other sources). Epidemiologists also communicate scientific findings through oral and poster presentations at scientific conferences and peer-reviewed publications.

Spontaneous reporting systems are the most commonly used pharmacovigilance method to generate signals on new or rare adverse events not discovered in clinical trials (see Chapter 10). However, there are several important limitations in interpreting spontaneous report data. Due to the lack of complete numerator (number of cases) data and the need to estimate the denominator (total number of patients actually exposed to the drug), it is not possible to determine the incidence of a particular event from spontaneous reports. Further evaluation of an apparent association between a drug and an adverse reaction usually requires postapproval epidemiologic studies.

Likewise, the nature of preapproval clinical trials often necessitates further safety evaluation through postapproval epidemiology. In addition to the limited sample size and length of follow-up of preapproval RCTs, with respect to drug safety an additional limitation of these studies lies in the common strict inclusion/exclusion criteria. Patients included in preapproval clinical studies may be the healthiest segment of that patient population. Special groups such as the elderly, pregnant women, or children are frequently excluded from trials. Patients in clinical trials also tend to be treated for well-defined indications, have limited and well-monitored concomitant drug use, and are closely followed for early signs and symptoms of adverse events which may be reversed with proper treatment.

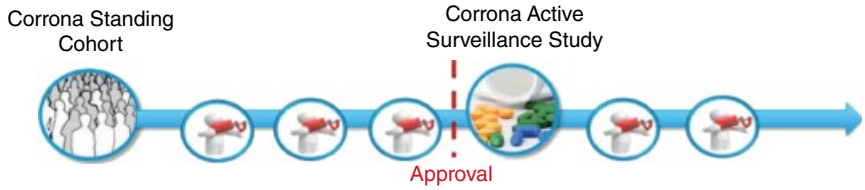
In contrast, once a drug is marketed, it is used in a real-world clinical context. Patients using the drug may have multiple co-morbidities for which they are being treated simultaneously. Patients may also be taking over-the-counter medications, "natural" remedies, or illicit drugs, unbeknown to the prescribing physician. The interactions of various drugs and treatments may result in a particular drug having a different safety profile in a postmarketing setting compared to the controlled premarketing environment. An example is the drug mibefradil, which was voluntarily withdrawn from the market by the manufacturer after less than a year as a result of new information about multiple potentially serious drug interactions. Adherence to medications also often differs between closely monitored trials and general postapproval use, as is the case with antihypertensives.

Because of the logistical complexity, high cost, and low external validity, large controlled trials have not been widely used for the postmarketing evaluation of drugs. Regulators and the medical community have communicated a desire for safety data from the populations that actually use the drugs in real-world clinical practice. This has led to a greater emphasis on the use of observational methods to understand the safety profile of new medications after they are marketed.

Tofacitinib Postapproval Epidemiology Strategy

As a condition for the approval of tofacitinib in the US, two observational postapproval safety studies were initiated: (i) an active surveillance study within the Corrona RA registry (ENCEPP/SDPP/5708) (see Figure 7.2), and (ii) a pregnancy outcome study within the OTIS registry (ENCEPP/SDPP/5703). Both studies were designed with the goal of characterizing the safety of tofacitinib within the real-world or clinical practice setting. In February 2018, both studies were in their fifth year.

Similarly, to assess the safety of tofacitinib in the EU postapproval setting, four five-year surveillance studies embedded within existing registries were proposed as a commitment to the EMA (i.e., ARTIS, BIOBADASER, BSRBR, and



- Broad patient population
- Tofa patients with 2 age- and gender-matched comparator groups
- ~5-year study with quarterly reports
- Complements the PMR RCT
- Analyses include evaluation of benefits and risks
- Starts with simple descriptive statistics, can include formal comparisons over time as data accumulate
- Multidisciplinary/cross-functional study team and internal steering committee

Figure 7.2 A closer look at the Xeljanz® (tofacitinib) postapproval Corrona active surveillance study. PMR, postmarketing requirements; RCT, randomized controlled trial.

RABBIT). The objective of the surveillance program is to evaluate any excess risk in the occurrence of known or potential adverse events, after accounting for confounding factors including disease severity and concomitant therapy. These studies, in addition to the aRMM-related work (i.e., HCP survey and drug utilization study) and ongoing pharmacovigilance (i.e., spontaneous reports, etc.), serve to address regulatory concerns regarding drug safety among European patients receiving care in the real world.

Drug Safety Evaluation in Special Populations

As already noted, most trials are conducted in populations that reflect a broad representation of the intended user to achieve regulatory approval. However, there are often populations of special interest, such as the elderly or children, who were omitted or insufficiently studied in these trials, necessitating additional research on these targeted populations.

In what follows we describe some additional examples of epidemiology studies in these special populations, and outline some of the unique challenges to and potential solutions for performing observational research in these groups.

Pregnancy and Birth Outcomes

The safety of a medicine when used during pregnancy is often unknown at the time of marketing approval (see also Chapter 22). Unless a medication is being developed specifically to treat a pregnancy-related condition or a vaccine is developed for maternal immunization, pregnant women are generally excluded from clinical trials for ethical reasons, due to potential risks to the developing fetus and newborn. In addition, most clinical trials that enroll women cease study of pregnant women upon detection of pregnancy. Thus, at the time of introduction to market, the effects of many medications on pregnancy are not well known, with the foundation of drug safety during pregnancy often resting largely on animal reproductive toxicology studies, whose extrapolation to humans is questionable. The paucity of data is potentially a serious concern for public health, particularly if the medication will be used by many women of

childbearing potential, since approximately half of all pregnancies in the US are unplanned [27] and more than half of all pregnant women in Western countries take medication during pregnancy [28]. While postmarketing spontaneous adverse event reporting of pregnancy outcomes may be helpful for identifying extremely rare outcomes associated with medication use during gestation, the limitations of these data are well established. Therefore, well-designed observational studies have become the primary way of generating evidence on the benefit and risk of medication use in pregnant women and their offspring.

In certain circumstances, registries are used to obtain information about the safety of new medications during pregnancy. A pregnancy exposure registry is typically prospective and observational, conducted to actively collect information about medication exposure during pregnancy and subsequent pregnancy outcome. The FDA and EMA issued guidelines regarding the circumstances when it is appropriate to establish a pregnancy registry [29,30]. The FDA's Office of Women's Health maintains an online list of pregnancy exposure registries [31], including some for HIV/AIDS medications, the human papillomavirus (HPV) and hepatitis B vaccines, and medications for cancers, depression, migraine, diabetes, and other conditions. Such registries differ from passive postmarketing surveillance systems, in that they collect data from women prior to knowledge of pregnancy outcomes, which has the effect of minimizing recall bias. The prospective nature of properly designed pregnancy registries also allows them to examine multiple pregnancy outcomes within a single study. Ideally, a pregnancy registry will allow for increased generalizability by being population based. It may aid the study of a cause–effect assessment between drug exposure and outcome by being prospective in nature; by collecting information on the timing of drug exposure, detailed treatment schedule, and dosing; by using standard and predefined definitions for pregnancy outcomes and malformations; and by recording these data in a systematic manner. The registry will ideally also follow offspring of medication-exposed women for a prolonged period after birth, to allow for detection of any delayed malformations in children who seem normal at birth.

Although pregnancy registries have advantages over passive surveillance methods, there are a number of major limitations that should be carefully considered. One of the major challenges of a pregnancy registry is a low level of enrollment. For example, the pharmaceutical company GlaxoSmithKline sponsored four international registries, but none of them reached the enrollment milestone of 1000 pregnancies during the first 10 years of data collection, considered by the European Medicines Agency Committee for Medical Products for Human Use to be necessary to be representative of widespread exposure [32,33]. The voluntary nature of enrollment can result in selection bias if women opting to enroll differ from those who do not, in terms of factors associated with the underlying risk of the outcome being studied. Pregnancy registries may also suffer from referral bias, with healthcare professionals being more or less likely to enroll women based on their disease severity. Sometimes more than one-third of pregnancies are then lost to follow-up, which makes the representativeness of the study population questionable [34]. Further, pregnancy registries often rely on self-reported medication exposures, which may result in misclassification of exposures, and often do not capture data on pregnancy losses (especially early losses). Another major challenge of a pregnancy registry is to establish an appropriate comparator group, especially when there is a possibility that the medical condition being treated may itself be associated with the outcome of interest, such as diabetes or epilepsy. Ideally, a pregnancy registry should allow for effects of the medication on pregnancy outcome to be distinguished from the effects of the

disease state warranting the treatment, if applicable, on pregnancy outcome. This can be achieved by enrolling two comparator groups: pregnant women who are disease free and not on the medication under study, and pregnant women with the disease who are not undergoing treatment or who are on a different treatment. In practice, however, it is usually not feasible to have comparator groups within pregnancy registries, because it is difficult to enroll pregnant women who are disease free or not using medication. Thus, in many cases, only pregnant women with the disease using the drug of interest, or other treatments for the disease, are enrolled and followed.

Given the limitations of pregnancy registries and the rapid development of large computerized healthcare databases, electronic healthcare databases are increasingly being used to evaluate the safety of medicine exposure during pregnancy, due to their ability to identify large populations in a timely, efficient manner. Using electronic healthcare databases for pregnancy studies can minimize selection bias by identifying all pregnant women who are exposed to a medication, and can avoid recall bias by using prescription data that are independently recorded by prescribers or claims. Internal comparator groups can be identified that represent the general population or women with the same underlying condition exposed to no medication or other medications within the class of interest. However, there is a lack of information on whether the women actually took the medicine and the precise timing of exposure. In addition, over-the-counter medication exposure is not captured in these databases.

There are three main types of electronic healthcare databases: national population-based registers in Nordic countries (Denmark, Sweden, Norway, Iceland, and Finland), electronic patient medical records, and administrative claims for reimbursement of medical treatment and prescriptions (see Chapters 11–14). Each type of data source has its own unique strengths and limitations, in addition to the general strengths and limitations of electronic healthcare databases already discussed. A key strength of national population-based registers in Nordic countries is the mandatory reporting of all live births and stillbirths within a country, which allows the capture of exposure and outcomes data from a representative sample of women, alleviating any concerns about the generalizability of study findings. These registers routinely collect information on gestational age at birth, birth weight, congenital malformations, mother's reproductive history, type of birth, delivery characteristics, and complications. Data on body mass index, smoking status, and alcohol intake are collected in most registers.

However, there are a few unique challenges that are not seen in other types of epidemiologic studies when using administrative claims databases to evaluate the safety of medication use during pregnancy. First, the information from the mother and that from the infant must be linked to each other. Researchers have used a variety of methods to establish the mother–infant linkage, such as unique family identification numbers included in the health plan enrollment data, date of birth and delivery date, or co-insurance information of the newborn, with reported linkage rates ranging from 63% to 88% [35–38]. Secondly, data on gestational age, an important variable for determining the timing of exposure during pregnancy, is not directly recorded in claims data. Pregnancy-related diagnosis and procedure codes can be used to estimate trimesters and weeks of gestation in administrative claims databases. A number of algorithms have been validated against the gestational age information on birth certificates or in medical records [39,40]. Claims data contain information on a number of confounding factors, such as maternal demographics, medical conditions that require medical attention, and concomitant medication exposure. However, information on mother's reproductive history,

body mass index, smoking status, and alcohol intake are typically not available or documented completely.

The large number of pregnancies captured within electronic healthcare databases and the fact that the data are routinely collected make them a valuable tool for evaluation of the benefit and risk of medicine use during pregnancy. Even with the large populations of electronic healthcare databases, a single data source may not have enough pregnancies or outcomes of interest, depending on the prevalence of the medical condition being studied and the frequency of use of the medication being prescribed. Therefore, pooling of multiple data sources for specific pregnancy studies is often needed to achieve sufficient statistical power. For example, national register data from five Nordic countries were pooled to assess whether maternal use of selective serotonin reuptake inhibitors increases the risk of persistent pulmonary hypertension in the newborn [41]; automated data from three healthcare claims databases and one electronic medical record database were used to evaluate topiramate use in pregnancy and the prevalence of oral clefts [42]; and recently, a planned study to examine the safety of Trumenba® vaccine exposure during pregnancy using electronic data from multiple healthcare systems in the US, all of which participate in the Sentinel distributed network [43].

Pediatrics

In the context of drug development, children are considered a special population (see also Chapter 22). This categorization is due to the unique physiologic characteristics of children: developing organ systems often result in different and unpredictable pharmacokinetic and pharmacodynamic profiles compared to adults, beyond standard adjustments for smaller body size and weight. The special population designation is also a result of the special ethical issues associated with testing unapproved substances in this vulnerable population, who cannot provide true informed consent.

There is drug development legislation specifically for pediatrics in both the EU and US. They differ slightly in their requirements and timing, but they both focus on early preparation of pediatric development plans and early submission (typically in Phase II) of these plans to regulatory authorities for review and approval. Similar to the EU RMP, the paediatric investigation plan (PIP) in Europe and the pediatric study plan (PSP) in the US require extensive information on the background epidemiology of the disease in children by age and other demographic and geographic subgroups; epidemiologists may play an important role in providing this information. Waiver requests, because of insufficient numbers of pediatric patients or because the disease does not exist in pediatrics, are another area that epidemiologists may support. The EU PIP process also requires that the adult RMP be expanded to include pediatric indications and to expand or conduct relevant safety studies, if applicable. In both regions, separate pediatric observational safety studies as conditions of approval are fairly common, given the lack of long-term safety data that is typically available in this population.

An important consequence of the US and EU pediatric legislation is that companies are now required to integrate pediatric planning much earlier in development than before. It also requires that epidemiologists become more familiar with general pediatric research issues, as well as pediatric population sources and research networks, since any patent-protected or patent-eligible drug or biologic with a European filing, regardless of the point in its life cycle, must submit a PIP. The regulatory changes in the US and general attention to active safety monitoring of vulnerable populations have also strengthened the need for postapproval observational safety studies in pediatrics.

Unlike drug development programs in adults, which typically encompass thousands of

individuals treated prior to approval, many pediatric drug development programs comprise a single pivotal clinical study of several hundred children. This small study size may be due to the lower incidence/prevalence of the disease in children, or the reluctance of parents to expose their children to experimental therapies. The small sample sizes drive the need for and increased focus on adequate long-term safety monitoring beyond traditional passive surveillance (i.e., spontaneous reports) in pediatrics, yet raise interesting methodologic challenges for epidemiologists studying the safety of drugs in children. One is frequently faced with assessing both a very rare exposure and a very rare outcome of interest. This conundrum necessitates observational approaches, because a traditional RCT cannot be feasibly conducted, but also may require creative approaches to design and data source selection that would not be necessary in adult populations. Utilizing existing registries or creating a new one, active surveillance programs using either primary or secondary data collection modalities, or collaborations with pediatric specialty research networks are just a few potential solutions.

Pediatric rheumatology, pediatric oncology, and many rare diseases represent common therapeutic areas that face these methodologic and ethical issues with several examples from which to draw [44–47]. In other instances, children are the primary target or vector for disease, as in many common infectious diseases. The primary methodologic limitation facing the epidemiologist may no longer be adequate sample size, but instead finding an unexposed comparator group, such as when studying a vaccine that is part of a universal vaccination campaign. More information on the nuances of vaccine safety evaluation is available in Chapter 20. Both situations highlight the need for novel approaches and methodologies to better support long-term safety monitoring of biopharmaceuticals in children.

Epidemiology in Evaluation of Risk Mitigation Interventions

Epidemiology not only plays an important role in evaluation of the drug safety profile pre- and postapproval, but, as noted earlier, also makes significant contributions to the evaluation of the effectiveness of risk mitigation intervention measures (see also Chapter 24). This component of biopharmaceutical risk management has grown considerably in the last decade, with the US, EU members, Taiwan, Egypt, Australia, and a number of other countries implementing legislation that supports risk mitigation interventions.

Under FDAAA, the FDA can require a sponsor to submit a proposed REMS as part of its initial application, if the FDA finds that a REMS is necessary to ensure the benefits of the drug or biologic product outweigh the risks [48]. The FDA may also require a REMS postapproval based on new safety information. The FDAAA has defined this as any information obtained during the initial review process, as a result of postapproval studies, or from spontaneous reports [48]. REMS are intended to be utilized to reduce known or hypothetical risks when traditional minimization approaches (i.e., the product label) are insufficient. These tools generally fall into three categories: enhanced education, that is patient labeling (including Medication Guides) or communication plans such as prescriber training programs; elements to assure safe use (ETASU), such as requiring documentation of laboratory tests before each prescription, or restricting distribution only to those who are certified prescribers; and an implementation system to monitor and evaluate ETASU. A critical addition to this legislation that was particularly relevant to epidemiologists within industry was the requirement to perform assessments of the effectiveness of these risk minimization tools and to submit these to the

FDA for review at prescribed time points, generally at 18 months, 3 years, and 7 years. The FDA draft guidances on REMS assessments and survey design were issued in January 2019, and it is unknown when these will be finalized. The EU has similar legislation to require sponsors to implement aRMMs where necessary to ensure that the benefits outweigh the risks, and a similar requirement to assess the effectiveness of the aRMMs, although without defined timelines. These aRMM programs and assessments are described in the EU RMP [4,5,49].

Epidemiologists play a critical role in the design and implementation of these assessments because of their expertise in observational study design, survey design, data analysis, and program evaluation. For example, using an automated healthcare or claims database, assessments may measure compliance with monitoring guidelines or whether a contraindicated population is prescribed the drug. Assessments may also examine the frequency of occurrence of an adverse event of interest before and after implementation of the risk minimization tool. Most commonly, however, assessments measure prescriber, pharmacist, or patient comprehension of risk information or self-reported adherence to risk minimization behaviors, and require the epidemiologist to craft cross-sectional surveys specific to each recipient, drug, and associated unique risk profile, since standardized or validated questionnaires that measure these concepts do not exist. An example of a comprehensive Tysabri® REMS program is shown in Box 7.1.

The implementation of the REMS and aRMM legislation has highlighted a number of difficulties. The mandated assessments may be difficult to achieve, or to achieve within the US legislative timelines, for many reasons: the need to develop and pilot knowledge/comprehension surveys unique to each drug subject to a REMS; the requirement to design, implement, and assess complex safe use programs; the scarcity of patients treated with the drug of interest; or difficulties in identifying such patients through automated channels. The fractured healthcare and prescription delivery system in the US and the wide variety of health systems, legal and privacy requirements, and attitudes to these programs and research participation across Europe present a barrier to the efficient distribution of educational materials, to the implementation of many safe use elements, and to the scientifically valid evaluation of these programs overall. Unfortunately, there is relatively little scholarly work published on how best to assess these burdensome but important risk mitigation programs, how best to define success, and, where necessary, how best to improve them [54–57]. Knowledge in these areas continues to mature as more companies and the regulatory agencies garner additional experience, and we expect that existing guidance [5] will evolve. Risk mitigation evaluation is thus still an emerging area for epidemiologists in industry, but one that complements our specialized training and expertise.

Tofacitinib Risk Mitigation Evaluation

The US REMS for tofacitinib was originally approved on November 6, 2012 and consisted of a Medication Guide, a communication plan, and a timetable for submission of REMS assessment (i.e., surveys at 18 months, 3 years, and 7 years). The FDA deemed the 18-month REMS epidemiologic assessment adequate, citing that the survey data demonstrated that patients understood the risks associated with therapy. Therefore, in accordance with Section 505-1(g)(4)(B) of the Food, Drug, and Cosmetic Act, the FDA determined that maintaining the Medication Guide as part of the approved labeling was sufficient to address safety-related concerns. As such, it was no longer necessary to include the Medication Guide as an element of the approved REMS, and the survey of patient knowledge and understanding was removed.

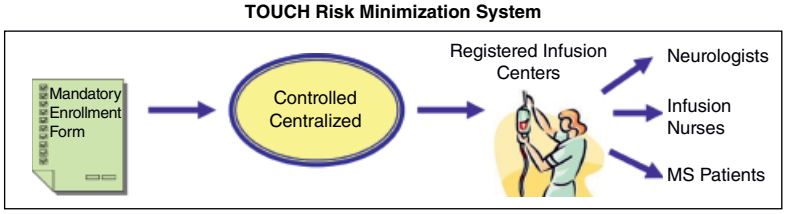
As a condition of approval within the EU, aRMMs were implemented within the EU,

Box 7.1 Tysabri® risk evaluation and mitigation strategies (REMS) case study

The FDA approved Tysabri (natalizumab), the first humanized monoclonal antibody for the treatment of relapsing multiple sclerosis (MS), via accelerated approval in November 2004. Approximately three months later, in February of 2005, Biogen/IDEC voluntarily suspended all sales and ongoing clinical studies due to two cases of progressive multifocal leukoencephalopathy (PML), one of which was fatal, in MS patients in long-term extension studies. Although no spontaneous reports had yet been reported to either the Sponsors or the FDA, the suspension was driven by the concern that the association between PML and natalizumab use was unclear, that PML is almost universally fatal, and that other patients may have undetected early-stage PML who would otherwise continue to receive the medication [50]. A third fatal case was identified in a Crohn's disease patient shortly thereafter [51].

At the time of suspension, little was known about the risk of PML in the general or MS population. In the general population, PML is extremely rare, and seldom occurs in immunocompetent individuals. It is estimated that 1–5% of AIDS patients may be diagnosed during their lifetime. PML also occurs in organ transplant recipients and cancer patients who have received immunosuppressive medications, but no cases in MS patients had previously been documented [50]. The Sponsors designed a comprehensive program to better understand the risk factors associated with PML development and developed a comprehensive RiskMAP program (later converted to a "Deemed REMS" under FDAAA) called the TOUCH™ (Tysabri Outreach: Unified Commitment to Health) Prescribing Program [52, 53]. The goals of TOUCH™ are as follows:

The program entails all REMS elements: a patient medication guide; additional physician education and accompanying assessments of knowledge and behavior; prescriber and patient attestation of risk understanding at enrollment; restricted distribution of the drug; and mandatory certification of physicians and infusion centers.



Based on this comprehensive program and a FDA Advisory Committee review in 2006, the clinical hold was lifted and natalizumab was reintroduced to the market in June of that year. Biogen/I DEC and regulatory agencies continue to closely monitor the incidence of PML potentially associated with natalizumab use (approximately 1 in 1000 in clinical trials) and communicate the updated safety information monthly to all stakeholders (e.g., neurologists, nurses, regulatory agencies). Additional research on risk factors and risk stratification, such as the impact of duration of use, the total number of infusions, and

the role of JC-virus infection, continue to be evaluated. The program has demonstrated a high degree of PML awareness and compliance with the requirements; most importantly, the fatality and disability rates appear lower than observed in clinical trials and the literature. Natalizumab was approved by the FDA in January 2008 for another indication, Crohn's disease, further supporting the effectiveness of the risk mitigation program [51, 53].

Key Points:

- Comprehensive risk assessment and mitigation plans (REMS) can preserve a positive benefit–risk balance in appropriate patient populations.
- Effective programs combine strict controls and tailored education, yet are dynamic (i.e., evolve as information becomes available and clinical practice and treatment options change over time).
- Transparent and frequent communication involving all stakeholders is critical, even if the safety profile is unknown or emerging.

consisting of an educational program intended to enhance the communication of the risks and risk minimization practices to patients and healthcare professionals (HCP). The elements of the aRMM include:

- Xeljanz® HCP brochure
- Xeljanz® HCP treatment initiation checklist
- Xeljanz® HCP treatment maintenance checklist
- Educational website
- Xeljanz® patient alert card

Under the purview of epidemiology, two studies were designed and implemented to evaluate the effectiveness of the aRMM. The studies are underway as of this writing and include a survey of HCPs and a drug utilization study.

Big Data and the Future of Pharmacoepidemiology

In recent years, the concept of "Big Data" has become more relevant to pharmacoepidemiology, with increases in the size and type of data available, as well as the computational capability to rapidly execute analyses across large datasets. Big Data is typically characterized by at least "3 Vs": velocity, variety, and volume; some commentators posit that veracity, variability, or value should be added to the list. However defined, Big Data and its associated analytic and machine learning techniques are having impacts on how pharmacoepidemiology is applied to drug development and safety assessment today. In the near future, we expect its use by industry to broaden across multiple functional activities, and its importance for decision making to increase within industry pharmacovigilance functions.

The greatest change in pharmacovigilance analytics being applied today, and the one most connected to the Big Data revolution, is wider use of observational data, as evidenced by pharmacoepidemiologic studies conducted across multiple databases and the development of large networks of observational databases of electronic healthcare records in North America, Europe, and Asia [58] (see also Chapter 25). There are now hundreds of existing longitudinal observational databases (LODs) available for secondary use in epidemiologic studies in North America, Europe, and Asia, from drug or outcome registries to transactional insurance claims databases and electronic medical record (EMR) databases [59], and increasingly these are being linked.

The best-known example of a LOD network for safety assessment is the US FDA's Sentinel system, initiated in pilot form in 2009 and consisting primarily of private transactional insurance claims data. The system was specifically designed to investigate potential safety concerns [60] to respond to the perceived weaknesses of a safety surveillance system reliant on spontaneous reporting. Sentinel, now in routine use at the FDA, conducts hundreds of assessments of products, conditions, and product–outcome pairs each year (Jeff Brown, personal communication). Analyses are conducted across a distributed network of data from 16 health plans (with additional datasets coming online over time), with currently over 220 million members and over 425 million person-years of data for analysis. Partners retain physical and operational control over their data by using a common data model (CDM). Standard, executable programs are then sent to each data partner to perform analyses or create analysis files for pooling summary data, which are then returned and compiled at a coordinating center. The network routinely uses standardized, simple queries that have as fast as a one-week turnaround from query initiation to result, a rapid analysis capability not seen previously on large-scale observational data. The distributed database, which is updated quarterly, has information on over 7 billion medical encounters and 6 billion outpatient pharmacy dispensations, and is growing at nearly 1 billion encounters a year (Jeff Brown, personal communication).

The Sentinel initiative is increasingly described as a component of a national evidence generation system [61]. In practice, this means exploring a broader use of this data network by connecting it to additional data types, including disease or drug registries, and/or other data networks, such as PCORNet, a network of EMR repositories [62], resulting in data systems encompassing more than half and up to two-thirds of the US population. These systems are envisaged as having value for research other than safety assessment, such as comparative effectiveness studies (see Chapter 26), pragmatic trials, or investigational trials in real-world settings. If the Sentinel system is truly a national resource, it also means developing governance and procedures for use by stakeholders other than the FDA, naturally with appropriate safeguards in place, to study important public health and safety research questions. The Reagan-Udall Foundation's Innovation in Medical Evidence Development and Surveillance (IMEDS) program was designed to develop a process for such access by stakeholders other than the FDA. In the first pilot conducted with IMEDS by Pfizer, policies and procedures were developed and subsequently tested by two use cases: the risk of venous thromboembolism associated with oral contraceptives; and the impact of an FDA labeling change on usage patterns of proton pump inhibitors [63–65].

Similar networks for safety surveillance have been developed around the world: AsPEN in Asia [66], CNODES in Canada [67], and several multinational European networks, such as the Innovative Medicines Initiative (IMI) PROTECT project [68], ARITMO [69], and SOS [70], while other networks, for instance OMOP and OHDSI, have primarily focused on methods testing and informatics tool development [71,72]. The Japanese regulatory authorities have also created a distributed network called MID-NET that links multiple Japanese hospitals and has a structure similar to that of Sentinel [73].

Many of these networks have developed CDM-based systems. CDM use for healthcare databases has clear benefits, but also limitations. For example, as ecosystems of tools grow to support efficient use of each CDM, this can lead to discordant results across systems [74]. A CDM is not a prerequisite for a data network. The IMI project PROTECT had a work package focused on pharmacoepidemiology that executed multiple study designs for six drug–outcome pairs across several European data sources using a

common protocol executed across data sources rather than a CDM. PROTECT demonstrated that through careful epidemiologic reasoning to produce a common protocol across multiple centers and countries, analyses conducted locally yielded generally concordant and predictably discordant results across databases [68]. Naturally, the amount of concordance varied by drug–outcome pair, study design, country, and database, but in a predictable manner.

As more observational data analyses are conducted, it is essential to ensure that studies are conducted only when there is a research question appropriate for observational study designs; techniques for confounding control are expected to be reasonably good at controlling for important confounders, particularly confounding by indication and severity; and regulatory and international scientific good practice guidelines are followed. Increased transparency in the conduct and reporting of studies may support better reproducibility and replicability. Recent guidance from the joint International Society of Pharmacoepidemiology–International Society for Pharmacoeconomics and Outcomes Research (ISPE–ISPOR) task force on "Real World Evidence in Health Care Decision Making" looks to provide guidance on the design and reporting of pharmacoepidemiologic analyses of longitudinal healthcare databases [75–78]. In the coming years, we expect further improvements in the quality and variety of clinical data available and linkages in these networks. For example, to study the prevalence of congenital malformations among infants exposed and not exposed to varenicline *in utero*, Danish and Swedish medical birth registries were used to identify live-born infants, then data on maternal varenicline use and congenital malformations in offspring were obtained by linkage to nationwide registries of dispensed prescriptions and hospital admissions [78]. Similarly, a planned study for meningitis B vaccination will look to examine the safety of Trumenba® vaccine exposure during pregnancy using electronic healthcare data and linked birth certificates from multiple healthcare systems in the US, all of which participate in the Sentinel distributed network [79]. Advances in these networks will also occur through data enrichment, such as linkage to clinical disease and drug registries or other primary data collection systems, or supplementing coded data with information obtained from the free text of medical records using natural language processing (NLP) or similar automated techniques [80].

The Big Data promises that have been most clearly fulfilled are the existence of these large networked data systems, coupled with the ability to gain insights into a study question within days rather than months or years as in the past [81]. With these advances, researchers have recently explored how these healthcare databases and systems might be used for exploratory assessment, rather than the usual signal evaluation approaches applied to these data. In this approach, the goal is to capture emerging and previously unsuspected signals; that is, hypothesis-free signal detection in healthcare databases [82]. There are limitations to the data, however, that currently hamper their usefulness in signal detection, including the lack of a learned reporter that suspects a medicine–adverse event relationship and is able to provide detailed medical information and a rationale for that suspicion; incomplete linkage of data from primary, specialist, and inpatient visits; and delays in updating and making available the data for research, which take up to a year for some databases. There is now a nascent literature on comparison of healthcare databases to spontaneous reports for signal detection, where there is cautious promise, at least for outcomes with high background event rates, which are difficult to capture as safety signals in spontaneous reporting systems [83]. For now, research suggests that, at best, this approach is likely to be complementary rather than to replace spontaneous reports.

Consumer wearable technology (such as fitness devices and smartphones) and "smart" digital technology (such as thermometers, or glucose or heart monitors) have the potential to supplement these approaches by providing better, more detailed health and behavioral information than that collected routinely in electronic healthcare databases. Some of the data are collected automatically, such as a phone's (and therefore individual's) location at a particular day and time. These data can then be linked to other information known about the location, for example the weather or air pollution levels. Most of the mobile data streams, though, are collected through "apps" that are intended to collect information about subjective experience and/or objective measures, such as heart rate or number of paces. These subjective data streams are potentially more representative and systematic than social media data streams, since they are able to prompt the user to enter data and provide data summaries that are useful or of interest to the user. For example, researchers in the UK are using smartphones and linked mobile data to study the relationship between weather patterns and RA symptom severity [84]. This study collects information about the severity of pain symptoms from an app and then links it to weather information based on the patient's location at time of data entry. To encourage participation and frequent data entry, the researchers have created the app so that it is relevant to patients: patients may view their individual symptom reporting over time as well as aggregate reporting trends for the entire study population. Elsewhere, computer games are used to collect data on reaction times to better understand disease progression [85], and smartphone apps are being explored as tools to collect safety information during research studies.

The use of consumer wearable technology for pharmacoepidemiologic research is in its infancy, although some researchers argue that the line between medical devices and consumer wearable technologies is already beginning to blur [86]. If large-volume data streams are being created that may be accessible in near real time and proximal temporally to a healthcare encounter or experience following the use of a medication, the promise for pharmacoepidemiology is great, particularly as these streams increasingly focus on medical and behavioral data (e.g., heart rate, personal and family medical history, smoking status, diet, alcohol consumption, and exercise patterns). Additionally, our analytic capabilities are being advanced, as these data streams and networks of sensors make it possible to examine relationships between data types previously unknown at this scale. To give three examples, not yet to our knowledge applied to drug safety research: studies demonstrating how noncontact visual images can be used to infer muscle activity and force [87], which one could anticipate would be of value in monitoring the progression of amyotrophic lateral sclerosis (ALS); video-recorded data can be used to allow more accurate health insights and therefore treatment in asthma patients [88]; and research into food-related object recognition [89] could potentially lead to objectively recorded dietary data being linked to electronic healthcare databases.

We anticipate that patients whose healthcare is complex, with an impact on their daily life, such as the chronically ill or those who are part of active and organized patient communities, will be early adopters of sharing their information for research purposes, despite the loss of privacy. This assumes of course that these data collection tools offer value to the individual patient and, when appropriate, their healthcare providers. Privacy, the perceived risk of misuse of data, and the regulatory considerations over tools that measure objective medical data (and therefore may be considered medical devices) are current hurdles to more widespread use. While it is impossible to predict the rapidity in uptake of these types of apps, we expect it will occur faster among younger generations of

healthcare providers and patients, who are arguably more comfortable with sharing data in this way and perhaps more likely to find it valuable.

As the amount of data of potential interest to companies has grown, so have computational advances for the analysis of these data sources. This has meant that traditional analytic approaches can be employed more rapidly, and/or extended in ways that were computationally intractable previously, such as large-scale high-dimensional propensity score matching and empirical Bayesian analysis approaches for shrinkage regression. Additionally, methods that would not have been practical in previous decades are now computationally viable analytic solutions on the large datasets of interest to pharmacoepidemiology. Examples of the latter are approaches for machine learning, artificial intelligence (AI), and cognitive computing. Currently, they appear most suited to making progress on questions requiring unsupervised pattern recognition, for example unexpected cluster detection in databases or predictive modeling [90]. Other technologies loosely connected to the Big Data movement may also have enormous impacts on pharmacoepidemiologic research strategies; for example, more widespread use of blockchain technologies may change the way we collect, structure, access, and thus analyze healthcare data.

Safeguarding patient safety and wellbeing is the ultimate purpose of these efforts to analyze new data streams and apply new analytic approaches. For this reason, as we consider the steady stream of new data and technology platforms of potential reliance to pharmacoepidemiology, we need to evaluate critically the impact of innovative data sources and techniques, and whether these should complement or replace existing approaches or are redundant, adding little or no value to current routine practice. Such evaluations will need to assess performance across data streams carefully, comparing their timeliness, effectiveness, and reliability for detecting emerging safety issues. Key to these evaluations is testing compared to appropriate, established external reference sets and transparent and measurable performance criteria. Objective and reproducible performance assessments are essential if such evaluations are intended to modify or enhance or replace components of current practice. There has been a recent focus in the field of pharmacovigilance on further developing the science of measuring the impact of pharmacovigilance activity.

Conclusions

Epidemiology makes a significant contribution to the development and marketing of safe and effective biopharmaceutical products worldwide. It facilitates the regulatory process and provides a rational basis for drug safety evaluation, particularly in the postapproval phase, and the evaluation of risk mitigation interventions. Like any other discipline, it must be properly understood and appropriately utilized. Industry has an opportunity to contribute to the development of the field and the responsibility to do so in a manner that expands resources while assuring scientific validity. With the passage of the 2007 FDAAA legislation and the 2010 EMA Regulation on Pharmacovigilance, the need for scientists with training and research experience in pharmacoepidemiology has never been greater. To best support drug safety evaluation, epidemiology strategies must (i) begin early in development, (ii) continue throughout the life cycle of the drug, (iii) evolve as new safety information becomes available, and (iv) be innovative, requiring epidemiologists to be aware of new methodologies and methods specific to the disease area. Epidemiologists within industry have an opportunity to build on the successes of the last 40 years by collaborating with academics, nonprofit organizations, and regulators to advance the methods of drug safety evaluation and risk management. In the near future, we expect the use and importance of Big Data to broaden within industry pharmacovigilance functions.

References

1 Institute of Medicine. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC: National Academies Press, 2007. https://doi.org/10.17226/11750

2 Food and Drug Administration. Postmarketing studies and clinical trials — Implementation of Section 505(O)(3) of the Federal Food, Drug, and Cosmetic Act. 2011. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf (accessed May 2019).

3 European Commission. The rules governing medicinal products in the European Union. Volume 9A: Guidelines on pharmacovigilance for medicinal products for human use. 2008. https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-9/pdf/vol9a_09-2008_en.pdf (accessed May 2019).

4 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies (Rev 3). 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf (accessed May 2019).

5 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module XVI – Risk minimisation measures: selection of tools and effectiveness indicators (Rev 2). 2017. https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/good-pharmacovigilance-practices (accessed May 2019).

6 Campbell UB, Walker AM, Gaffney M, *et al*. Acute nonarteritic anterior ischemic optic neuropathy and exposure to phosphodiesterase type 5 inhibitors. *J Sex Med* 2015; **12**(1): 139–51.

7 Flahavan EM, Li H, Gupte-Singh K, *et al*. Prospective case-crossover study investigating the possible association between nonarteritic anterior ischemic optic neuropathy and phosphodiesterase type 5 inhibitor exposure. *Urology* 2017; **Jul**(105): 76–84.

8 Dodd CN, Romio SA, Black S, *et al*.; Global H1N1 GBS Consortium. International collaboration to assess the risk of Guillain Barré Syndrome following Influenza A (H1N1) 2009 monovalent vaccines. *Vaccine* 2013; **31**(40): 4448–58.

9 Gault N, Castañeda-Sanabria J, De Rycke Y, Guillo S, Foulon S, Tubach F. Self-controlled designs in pharmacoepidemiology involving electronic healthcare databases: a systematic review. *BMC Med Res Methodol* 2017; **17**(1): 25.

10 Glynn RJ, Monane M, Gurwitz JH, Choodnovskiy I, Avorn J. Agreement between drug treatment data and a discharge diagnosis of diabetes mellitus in the elderly. *Am J Epidemiol* 1999; **149**(6): 541–9.

11 Melfi CA, Croghan TW. Use of claims data for research on treatment and outcomes of depression care. *Med Care* 1999; **37**(4): AS77–80.

12 Lewis JD, Brensinger C, Bilker WB, Strom BL. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf* 2002; **11**(3): 211–18.

13 Metlay JP, Hardy C, Strom BL. Agreement between patient self-report and a Veterans Affairs national pharmacy database for identifying recent exposures to antibiotics. *Pharmacoepidemiol Drug Saf* 2003; **12**(1): 9–15.

14 Hennessy S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003; **12**(2): 103–11.

15 Eng P, Seeger JD, Loughlin J, *et al*. Serum potassium monitoring for users of ethinyl estradiol/drospirenone taking medications predisposing to hyperkalemia: physician

compliance and survey of knowledge and attitudes. *Contraception* 2007; **75**(2): 101–7.

16 Eng P, Johannes CB, Chaing CC, West WA, Seeger JD, Walker AM. Survey of disability, medication use, and medication satisfaction among members with rheumatoid arthritis of a large health care plan database. *Pharmacoepidemiol Drug Saf* 2005; **14**: S5.

17 International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**(2): 200–8.

18 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 6). 2017. http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE_Rev6.pdf (accessed May 2019).

19 Food and Drug Administration. Guidance for industry and FDA staff: best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. 2013. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-conducting-and-reporting-pharmacoepidemiologic-safety-studies-using-electronic (accessed May 2019).

20 Faich GA, Lawson DH, Tilson HH, Walker AM. Clinical trials are not enough: drug development and pharmacoepidemiology. *J Clin Res Drug Dev* 1987; **1**: 75–8.

21 Tennis P, Cole TB, Annegers JF, Leestma JE, McNutt M, Rajput A. Cohort study of incidence of sudden unexplained death in persons with seizure disorder treated with antiepileptic drugs in Saskatchewan, Canada. *Epilepsia* 1995; **36**(1): 29–36.

22 Derby LE, Tennis P, Jick H. Sudden unexplained death among subjects with refractory epilepsy. *Epilepsia* 1996; **37**(10): 931–5.

23 Leestma JE, Annegers JF, Brodie MJ, *et al*. Sudden unexplained death in epilepsy: observations from a large clinical development program. *Epilepsia* 1997; **38**(1): 47–55.

24 Mussen F, Salek S, Walker S. A quantitative approach to benefit–risk assessment of medicines – part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf* 2007; **16**: S2–S15.

25 Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Stat Med* 2011; **30**(12): 1419–28.

26 Coplan P, Noel R, Levitan B, Ferguson J, Mussen F. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit–risk balance of medicines. *Clin Pharmacol Ther* 2011; **89**: 312–15.

27 Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008–2011. *N Engl J Med* 2016; **374**(9): 843–52.

28 Daw JR, Hanley GE, Greyson DL, Morgan SG. Prescription drug use during pregnancy in developed countries: a systematic review. *Pharmacoepidemiol Drug Saf* 2011; **20**(9): 895–902.

29 Food and Drug Administration. Guidance for industry: establishing pregnancy exposure registries. 2002, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071639.pdf (accessed May 2019).

30 European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). Guideline on the exposure to medicinal products during pregnancy: need for post-authorisation data. 2005. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/11/WC500011303.pdf (accessed May 2019).

31 Mastroianni AC, Faden R, Federmen D. *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Trials*. Washington, DC: National Academy of Science Press, 1994.

32 European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). Guideline on risk assessment of medicinal products on human reproduction and lactation: from data to labelling. 2008. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-risk-assessment-medicinal-products-human-reproduction-lactation-data-labelling_en.pdf (accessed May 2019).

33 Charlton RA, Cunnington MC, de Vries CS, Weil JG. Data resources for investigating drug exposure during pregnancy and associated outcomes. *Drug Saf* 2008; **31**(1): 39–51.

34 Kendle International Inc. Buproprion Pregnancy Registry: Final report 1, September 1997 through 31 March 2008. 2008. http://pregnancyregistry.gsk.com/documents/bup_report_final_2008.pdf (accessed May 2019).

35 Cole JA, Modell JG, Haight BR, Cosmatos IS, Stoler JM, Walker AM. Bupropion in pregnancy and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf* 2007; **16**(5): 474–84.

36 Davis RL, Rubanowice D, McPhillips H, *et al*. Risks of congenital malformations and perinatal events among infants exposed to antidepressant medications during pregnancy. *Pharmacoepidemiol Drug Saf* 2007; **16**(10): 1086–94.

37 Layton JB, Butler AM, Li D, *et al*. Prenatal Tdap immunization and risk of maternal and newborn adverse events. *Vaccine* 2017; **35**(33): 4072–8.

38 Garbe E, Suling M, Kloss S, Lindemann C, Schmid U. Linkage of mother–baby pairs in the German Pharmacoepidemiological Research Database. *Pharmacoepidemiol Drug Saf* 2011; **20**(3): 258–64.

39 Li Q, Andrade SE, Cooper WO, *et al*. Validation of an algorithm to estimate gestational age in electronic health plan databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(5): 524–32.

40 Margulis AV, Palmsten K, Andrade SE, *et al*. Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. *Pharmacoepidemiol Drug Saf* 2015; **24**(4): 335–42.

41 Kieler H, Artama M, Engeland A, *et al*. Selective serotonin reuptake inhibitors during pregnancy and risk of persistent pulmonary hypertension in the newborn: population based cohort study from the five Nordic countries. *BMJ* 2012; **344**.

42 Mines D, Tennis P, Curkendall SM, *et al*. Topiramate use in pregnancy and the birth prevalence of oral clefts. *Pharmacoepidemiol Drug Saf* 2014; **23**(10): 1017–25.

43 Panozzo CA, Purcell B, Andrade S, *et al*. Safety of Trumenba vaccine among pregnant women in the United States: planning and design of a large-scale multi-site observational study. *Pharmacoepidemiol Drug Saf* 2017; **26**(S2): 410.

44 O'Leary M, Krailo M, Anderson JR, Reaman GH. Progress in childhood cancer: 50 years of research collaboration, a report from the Children's Oncology Group. *Semin Oncol* 2008; **35**(5): 484–93.

45 Ringold S, Hendrickson A, Abramson L, *et al*. A novel method to collect medication adverse events in juvenile arthritis: results from the Childhood Arthritis and Rheumatology Research Alliance enhanced drug safety surveillance project (EDSSP). *Arthrit Care Res* 2015; **67**(4): 529–34.

46 Rowe SM, Borowitz DS, Burns JL, *et al*. Progress in cystic fibrosis and the CF Therapeutics Development Network. *Thorax* 2012; **67**(10): 882–90.

47 Sobel RE, Lovell DJ, Brunner HI, *et al*.; for the PRCSG. Safety of celecoxib and non-selective

non-steroidal anti-inflammatory drugs in juvenile idiopathic arthritis: results of the Phase 4 Registry. *Pediatr Rheumatol* 2014; **12**: 29.

48 FDA. REMS: FDA's application of statutory factors in determining when a REMS is necessary: guidance for industry. 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rems-fdas-application-statutory-factors-determining-when-rems-necessary-guidance-industry (accessed May 2019).

49 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module V – Risk management systems (Rev 2). 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-module-v-risk-management-systems-rev-2_en.pdf (accessed May 2019).

50 Holmes LB, Wyszynski DF, Lieberman E. The AED (antiepileptic drug) pregnancy registry: a 6-year experience. *Arch Neurol* 2004; **61**: 673–8.

51 Tomson T, Battino D, Craig J, *et al*. Pregnancy registries: differences, similarities, and possible harmonization. *Epilepsia* 2010; **51**: 909–15.

52 Smith MY, Sobel RE, Wallace CA. Monitoring the long-term safety of therapies for children with juvenile idiopathic arthritis: time for a consolidated patient registry. *Arthritis Care Res (Hoboken)* 2010; **62**: 800–4.

53 European Medicines Agency. Long-term pharmacovigilance for adverse effects in childhood arthritis focusing on immune modulatory drugs. http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2010/06/WC500091508.pdf (accessed June 2011).

54 Department of Health and Human Services. Office of Inspector General: FDA lacks comprehensive data to determine whether REMS improve drug safety. 2013. https://oig.hhs.gov/oei/reports/oei-04-11-00510.asp (accessed May 2019).

55 Gridchyna I, Cloutier AM, Nkeng L, Craig C, Frise S, Moride Y. Methodological gaps in the assessment of risk minimization interventions: a systematic review. *Pharmacoepidemiol Drug Saf* 2014; **23**(6): 572–9.

56 Kesselheim AS, Donneyong M, Dal Pan GJ, *et al*. Changes in prescribing and healthcare resource utilization after FDA Drug Safety Communications involving zolpidem-containing medications. *Pharmacoepidemiol Drug Saf* 2017; **26**(6): 712–21.

57 Nyeland ME, Laursen M, Callréus T. Evaluating the effectiveness of risk minimisation measures: the application of a conceptual framework to Danish real-world dabigatran data. *Pharmacoepidemiol Drug Saf* 2017; **26**(6): 607–14.

58 Trifirò G, Coloma PM, Rijnbeek PR et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med* 2014; **275**(6): 551–61.

59 Hennessy S. Use of health care databases in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006; **98**(3): 311–13.

60 Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network: improving the evidence of medical-product safety. *N Engl J Med* 2009; **361**(7): 645–7.

61 Califf RM, Robb MA, Bindman AB, *et al*. Transforming evidence generation to support health and health care decisions. *N Engl J Med* 2016; **375**(24): 2395–400.

62 Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014; **21**(4): 578–82.

63 Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's Sentinel initiative: a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther* 2016; **99**(3): 265–8.

64 Bate A, Sobell RE, Marshall J, *et al*. Oral contraceptives and VTE across the Sentinel

data network: an IMEDS evaluation pilot assessment. *Pharmacoepidemiol Drug Saf* 2016; **25**(S3).

65 Sobel RE, Bate A, Marshall J, *et al*. Risk minimization evaluation in a distributed data network: an IMEDS evaluation pilot assessment of the 2010 class label change for proton pump inhibitors. *Pharmacoepidemiol Drug Saf* 2016; **25**(S3).

66 AsPEN collaborators; Andersen M, Bergman U, Choi NK, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multinational collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; **22**(7): 700–4.

67 Suissa S, Henry D, Caetano P, *et al*. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med* 2012; **6**(4): e134–e40.

68 Klungel OH, Kurz X, de Groot MC, *et al*. Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 156–65.

69 Mor A, Frøslev T, Thomsen RW, *et al*. Antibiotic use varies substantially among adults: a cross-national study from five European countries in the ARITMO project. *Infection* 2015; **43**(4): 453–72.

70 Valkhoff VE, Schade R, 't Jong GW, *et al*. Population-based analysis of non-steroidal anti-inflammatory drug use among children in four European countries in the SOS project: what size of data platforms and which study designs do we need to assess safety issues? *BMC Pediatr* 2013; **13**: 192.

71 Hripcsak G, Duke JD, Shah NH, *et al*. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; **216**: 574–8.

72 Stang PE, Ryan PB, Racoosin JA, *et al*. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med* 2010; **153**(9): 600–6.

73 Ishiguro C, Takeuchi Y, Uyama Y, Tawaragi T. The MIHARI project: establishing a new framework for pharmacoepidemiological drug safety assessments by the Pharmaceuticals and Medical Devices Agency of Japan. *Pharmacoepidemiol Drug Saf* 2016; **25**: 854–9.

74 Xu Y, Zhou X, Suehs BT, *et al*. A comparative assessment of observational medical outcomes partnership and mini-Sentinel common data models and analytics: implications for active drug safety surveillance. *Drug Saf* 2015; **38**(8): 749–65.

75 Bate A. Guidance to reinforce the credibility of health care database studies and ensure their appropriate impact. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1013–17.

76 Berger ML, Sox H, Willke R. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the Joint ISPOR-ISPE Special Task Force on real-world evidence in healthcare decision making. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1033–9.

77 Wang SV, Schneeweiss, S. On behalf of the joint ISPE-ISPOR Special Task Force on real-world evidence in health care decision making: reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1018–32.

78 Olsen M, Petronis KR, Frøslev T, *et al*. Maternal use of varenicline and risk of congenital malformations. *Pharmacoepidemiol Drug Saf* 2015; **24**(S1): 244.

79 Panozzo CA, Purcell B, Andrade S, *et al*. Safety of Trumenba vaccine among pregnant women in the United States: planning and design of a large-scale multi-site observational study. *Pharmacoepidemiol Drug Saf* 2017; **26**(S2): 410.

80 Walker AM, Zhou X, Ananthakrishnan AN, *et al*. Computer-assisted expert case definition in electronic health records. *Int J Med Inform* 2016; **86**: 62–70.

81 Curtis LH, Weiner MG, Boudreau DM, *et al*. Design considerations, architecture, and use of the Mini-Sentinel distributed data sysem. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 23–31.

82 Noren GN, Sundberg R, Bate A, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov* 2010; **20**(3): 361–87.

83 Pacurariu AC, Straus SM, Trifirò G, *et al*. Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. *Drug Saf* 2015; **38**(12): 1201–10.

84 Reade S, Spencer K, Sergeant JC, *et a*l. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR Mhealth Uhealth* 2017; **5**(3): e37.

85 Strickland E. Akili: diagnosing Alzheimer's with a game. *IEEE Spectrum* 2014; **51**(5): 25–6.

86 Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. *PLoS Med* 2016; **13**(2): e1001953.

87 Sagawa R, Yoshyyasu Y, Alspach A, Ayusawa K, Yamane K, Hilton A. Analyzing muscle activity and force with skin shape captures by non-contact visual sensor. *Pacific-Rim Symposium on Image and Video Technology*. New York: Springer, 2015.

88 Rich M, Lamola S, Amory C, Schneider L. Asthma in life context: Video Intervention/ Prevention Assessment (VIA). *Pediatrics* 2000; **105**(3 Pt 1): 469–77.

89 Bolanos M, Garolera M, Radeva P. Active labeling application applied to food-related object recognition. *Proceedings of the 5th International Workshop on Multimedia for Cooking & Eating Activities*. New York: ACM, 2013.

90 Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017; **24**(1): 198–208.

8

The Role of Pharmacoepidemiology in Regulatory Agencies

Gerald J. Dal Pan[1], June Raine[2], and Shinobu Uzu[3]

[1] *Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA*
[2] *Medicines and Healthcare Products Regulatory Agency, London, UK*
[3] *Pharmaceuticals and Medical Devices Agency, Tokyo, Japan*

The regulation of pharmaceuticals aims at ensuring that the public has access to medicines that are effective, acceptably safe, and of high quality. A wide range of regulatory activities spans the entire life cycle of a medicine, involving laboratory-based understanding of pharmacologic action, animal testing, providing scientific and regulatory input to drug development programs, protection of human subjects during clinical trials, assuring the integrity of the manufacturing process, reviewing the dossier to support product approval or licensure, monitoring the safety of medicines after they enter the market, and many other activities. These regulatory activities are firmly rooted in science, have a strong public health focus, and are executed within a legal and regulatory framework.

The Scope of Pharmacoepidemiology throughout the Medicinal Product Life Cycle

Assessing the Need for Medicines

Pharmacoepidemiology, along with other areas of medical epidemiology, can be used in drug development long before a medicine is licensed or even tested in humans. Pharmacoepidemiologic approaches can be used to examine patterns of utilization of existing disease treatments, in order to identify and characterize disease populations and subpopulations for which unmet medical needs exist. In some cases, there may be no available therapies. In other cases, available therapy may be ineffective for or poorly tolerated by

The views expressed herein are those of the authors, and not necessarily of the US Food and Drug Administration, the European Medicines Agency, or the Japanese Pharmaceuticals and Medical Devices Agency.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

certain patients. In these cases, pharmacoepidemiologic approaches can be used to characterize patients who experience a suboptimal response to the medicine, and thus to define the target population for a drug development program. For example, population-based databases can be used to characterize the frequency and distribution of characteristics of patients with a specific disease, so that relevant populations can be included in the developmental clinical trials. Healthcare databases can be used to estimate the frequency of co-morbid conditions in the setting of the specific underlying disease to be treated, so that relevant background rates can be derived to place in context potential adverse events that arise during development. This is especially useful for clinical events that are seen more frequently in patients with the disease for which the new treatment is being tested, but which could also represent an adverse drug reaction. This situation, known as confounding by indication (see also Chapters 3, 33, and 43), is a well-known methodologic problem in observational pharmacoepidemiologic studies, but can also complicate the interpretation of adverse events in clinical trials, especially if the trial is not designed or powered to analyze these events. In these situations, careful understanding of background rates can be important.

Orphan Drugs

In the last decade, there has been substantial activity and progress in the development of drugs for rare diseases [1]. Orphan drug programs are designed to provide incentives to pharmaceutical manufacturers that develop medicines for rare conditions, known as "orphan drugs." In the US, an orphan drug designation is given to a drug or biologic that has shown promise as a therapy intended to treat a disease affecting fewer than 200 000 persons in the country [2]. In Japan, orphan designation is granted for drugs or medical devices if they are intended for use in fewer than 50 000 patients in the nation and for which there is a high medical need [3]. In the European Union (EU), a prevalence rate of 5 per 10 000 persons in the EU is used [4]. When all rare diseases are taken together, their public health impact is significant; approximately 25 million people in North America are affected by these diseases [5].

Medical epidemiology is central to the designation of a product as an orphan drug, as determination of prevalence is the basis for such a designation. Data sources for determining prevalence can include administrative healthcare databases, electronic medical record systems, registries, and surveys. In many cases, combining data from multiple sources will be necessary. In most cases, data from these sources, even when combined, will not cover the entire jurisdiction for which the orphan designation applies. Thus, some form of extrapolation must be performed to determine if the relevant population prevalence has been exceeded. Most orphan drug designations are for diseases or conditions whose prevalence is much lower than the 200 000 prevalence threshold in the US. A review of 25 years' experience with the orphan drug program in the US, covering 1892 orphan designations, found that the median prevalence was 39 000; the most common patient prevalence was 10 000 or fewer patients, with relatively few prevalence rates near the 200 000 threshold. For estimates of population prevalence near the threshold, care must be taken to ensure that the most rigorous methods have been used to estimate the population prevalence of a rare disease. The closer the estimated prevalence is to the threshold, the greater the precision needed to characterize the prevalence.

Planning Drug Development Programs

Despite the availability of an increasing number of medicines, there remain a substantial number of unmet medical needs. Advances in understanding the molecular pathogenesis of

cancers, rare diseases, and infectious diseases have led to a rapid rise in the number of drug development programs targeting these conditions. At the same time, the aging population across the globe has led to a need for improved treatments for widespread diseases such as diabetes, hypertension, ischemic heart disease, chronic obstructive pulmonary disease, Alzheimer disease, other neurodegenerative disorders, and many others. The continuing emergence of antibacterial resistance and the threat of new viral illnesses prompt the need for new antimicrobial agents. Infections with *Mycobacterium tuberculosis*, *Plasmodium falciparum*, human immunodeficiency virus (HIV), endemic parasitic diseases, and other agents contribute substantially to the global burden of disease and require new treatments [6].

Regulatory agencies have responded to this demand with a variety of regulatory programs and pathways designed to promote efficient development of medicines and to reduce drug development time so that these unfulfilled medical needs can be met. Some of these programs seek to optimize drug development by providing timely consultation between the regulator and the company developing a drug to clarify scientific requirements; other programs allow clinical development to be shortened by allowing the use of surrogate markers rather than clinical markers. In the US, the breakthrough therapy designation [7] allows the Food and Drug Administration (FDA) to work closely with drug sponsors to plan a development program that will efficiently generate evidence of effectiveness and safety for drugs that meet two general criteria: (i) the drug must be for a disease or condition that is serious or life-threatening; and (ii) there is preliminary evidence that the drug may demonstrate a substantial improvement over existing therapy. When breakthrough designation is granted, the agency commits to provide intensive advice on drug development, involve senior managers throughout development, review portions of the application before the complete application is submitted (a "rolling review"), and take other actions to expedite review, as necessary [8]. The European Medicines Agency (EMA) launched the PRIME (PRIority MEdicines) initiative to provide support for the development of medicines that show a potential to benefit patients with an unmet medical need. For drugs selected for PRIME, the EMA provides guidance on the overall development program as well as scientific advice at key development milestones, assigns rapporteurs from relevant EMA committees and a dedicated point of contact, and determines whether the marketing application may be eligible for accelerated assessment [9]. In Japan, the SAKIGAKE review program was introduced as part of the "Japan Revitalization Strategy" to improve access to medicines [10]. *Sakigake* is a Japanese word meaning "frontrunner" or "pioneer." The SAKIGAKE review program applies to medicines that meet four general criteria: (i) a mechanism of action different from that of other drugs; (ii) the target disease is serious and life threatening, or causes chronic disabling symptoms, for which there is no cure; (iii) there is no approved product or product anticipated to be markedly more effective than existing treatments; and (iv) there is the intent to have early development and initial approval in Japan. The SAKIGAKE designation allows for priority consultations with the Japanese Pharmaceutical and Medical Devices Agency (PMDA), rolling review, priority review, and other features that enhance sponsors' interactions with the PMDA. In each of these situations, pharmacoepidemiologic analyses can aid in the comparison of new treatments to existing treatments, especially when data on existing treatments are derived from clinical experience and not from formal clinical trials.

The goal of a drug development program is to demonstrate that a medicine has a beneficial and meaningful effect on a clinically important outcome, generally a measure of how the patient feels, functions, or survives. Clinical

trials whose primary endpoint is a direct measure of a clinically important outcome may be very long and delay patients' access to effective therapies. To allow patient access as rapidly as is feasible, and to ensure that definitive evidence of effectiveness is obtained, an alternative approach allows marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has a beneficial effect on a surrogate endpoint. A surrogate endpoint is an outcome measure that is used in place of a direct measure of a clinically meaningful outcome when the effect of treatment on the surrogate endpoint is expected to reflect changes in the clinically meaningful outcome [11]. In the context of drug development, a validated surrogate endpoint is one for which evidence exists that the effect of treatment on the surrogate endpoint predicts the effect of treatment on the clinical outcome of interest. For example, systolic blood pressure is used as a surrogate endpoint in clinical trials of antihypertensive agents, because it predicts the risk of occurrence of stroke. Similarly, HIV viral load is used as a surrogate endpoint in clinical trials of antiretroviral agents, because it predicts the development of an acquired immunodeficiency syndrome diagnosis [12]. Validated surrogate markers are widely employed to support approval of medicines.

There are, however, many serious and life-threatening conditions for which there are no validated surrogate markers, yet there is still an urgent need to bring effective therapies to patients in a timely way. For this latter situation, the concept of "accelerated approval" has been developed. Under this framework, the US FDA may grant approval to a medicine intended to treat a serious or life-threatening disease based on an unvalidated surrogate endpoint that is reasonably likely, depending on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity [13]. In these cases, postmarketing studies must be conducted to demonstrate the actual clinical benefit of the medicine [14].

A key regulatory tool in the EU to fulfill unmet medical needs is the conditional marketing authorization, which has reduced data requirements linked to a one-year, time-limited authorization, where the authorization's renewal is linked to further data submission [15]. Under the applicable regulations, manufacturers must study the drug further once it is approved, to verify and describe its clinical benefit, where there is uncertainty about the relationship of the surrogate endpoint to clinical benefit, or of the observed clinical benefit to ultimate outcome. At the time of approval, postmarketing studies would usually already be underway.

In Japan, the Pharmaceuticals, Medical Devices and Other Therapeutics Products (PMD) Act established a system of conditional and time-limited approval for regenerative medicines based on probable benefit from early clinical trials [16]. After obtaining such approval, the marketing authorization holder is required to submit a standard marketing application with additional data on safety and efficacy. Similarly, from 2017 a conditional early approval program applies to drugs offering high efficacy and clinical usefulness in the treatment of serious diseases, drugs for which conducting confirmatory studies is impracticable, and other designated drugs. One prerequisite for approval will be a commitment to complete postmarketing studies as necessary in order to reconfirm product safety and efficacy.

Understanding the relationship between a surrogate endpoint and a clinically relevant endpoint, as well as validation of the surrogate endpoint, is an opportunity for pharmacoepidemiologists to contribute to drug development. Pharmacoepidemiologists can use principles of epidemiology to distinguish simple correlation between a potential endpoint and a clinically meaningful outcome, on the one hand, from a true surrogate marker. For example, a marker of

disease status used in natural history studies may not be an adequate surrogate endpoint in a clinical trial, because it is not related to the disease mechanisms that give rise to symptoms, morbidity, and mortality [17].

Preapproval Review of Clinical Safety Data

While the traditional role of pharmacoepidemiology, from a regulatory standpoint, has been the assessment of the safety of medicines in the postlicensing period, pharmacoepidemiology can play an important role during the prelicensing review of safety data. The limitations of prelicensing clinical trials in defining the full scope of adverse drug reactions are mainly related to the fact that clinical trials are relatively small in size, compared to the population of patients that will ultimately take the medicine once it is marketed. Patients who participate in clinical trials may have fewer co-morbidities and take fewer concomitant medications than those treated in actual practice. Prelicensing clinical trials generally provide relatively little data, or no data at all, in certain populations such as children, the elderly, and pregnant women, or at-risk groups such as immunosuppressed patients. These groups, however, are treated with the medicine in the course of clinical practice once it is licensed.

The analytic methods of clinical trials are best suited for data arising from randomized, controlled, comparative trials. Many clinical trials of medicines intended for chronic or long-term use, including those trials in preapproval drug development programs, may have single-arm, open-label extensions after participants have completed the randomized portion of the trial. For data generated from this portion of the clinical trial, the techniques of observational pharmacoepidemiology may be appropriate. In addition to tallying the frequencies of specific adverse events, data from long-term extension studies can be examined to characterize patterns of adverse event onset over time. If appropriate, analyses based on person-time can be performed. In this setting, the interpretations of adverse events must take into account the prior treatment received during the randomized portion of the trial, the duration of treatment, the underlying frequency of medical outcomes in the population with the disease being treated, and other factors. Pharmacoepidemiology can inform this approach.

Planning for Postapproval Studies

At the time a medicine is approved, there are uncertainties and unknowns regarding its safety profile. In many cases, the nature of the safety issues that will unfold postapproval cannot be predicted at the time the product is brought to market. In some cases, however, a careful review of the clinical data at the time of approval can lead to a proactive approach to obtaining more safety information.

Pharmacoepidemiology can play an important role in several specific situations. First, drug development programs based on the use of unvalidated surrogate markers, as described earlier, generally require postmarketing studies to demonstrate definitively the clinical effectiveness of the product. In these situations, pharmacoepidemiologists can be involved in studies assessing the validity of the surrogate marker.

Secondly, pharmacoepidemiologists can be involved in the design and interpretation of postmarketing studies designed to assess the impact of new formulations of medicines developed to have a more favorable safety profile than earlier versions. For example, the widespread abuse of opioid-containing drug products has generated interest in the development of abuse-deterrent formulations of these products. While a variety of different physicochemical and pharmacologic mechanisms can confer the abuse-deterrent property, the true public health impact of the reformulation can be assessed only through formal epidemiologic analysis when the product is in actual use [18].

Thirdly, pharmacoepidemiologists can be involved in planning postmarketing studies when safety signals are detected prior to approval. An example of a proactive approach is the strategy the US FDA has developed to require sponsors of antidiabetic agents to characterize as fully as possible the cardiovascular risks of these medicines [19]. The strategy starts prior to approval, when data from clinical trials are examined to determine the cardiovascular risk of the new medicine to that of comparative agents. A relative risk estimate is calculated. If the upper limit of the 95% confidence limit of this estimate exceeds 1.8, the product will require a large cardiovascular outcomes clinical trial prior to approval. If the upper bound of the 95% confidence limit falls between 1.3 and 1.8, the product can be marketed, provided of course that all other criteria for approval are met, and the manufacturer will be required to conduct a postapproval clinical trial to determine the frequency of adverse cardiovascular outcomes relative to other antidiabetic agents. If the upper limit of the 95% confidence interval is below 1.3, and the product otherwise qualifies for approval, no further cardiovascular study is needed. This strategy provides a tiered approach, spanning the pre- and postapproval periods, to assessing the cardiovascular risks of antidiabetic agents, and accounts for the level of uncertainty in the preapproval data.

Monitoring Postapproval Safety

For the regulator, the postmarketing assessment of the safety of medicines involves both a proactive approach and, of necessity, a reactive approach. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) has developed a useful and practical framework that summarizes the known safety issues of a product and can form the basis of ongoing monitoring and, as needed, specific studies [20]. The ICH framework characterizes important identified risks, important potential risks, and important missing information. This framework allows pharmacoepidemiologists and others to devise proactive strategies to design observational studies or clinical trials to address unanswered questions about the safety profile of a medicine. The approach to studying the cardiovascular risk of antidiabetic agents, already noted, is an example of a proactive step taken at the time of approval. However, the identification of knowledge gaps can occur at any time in the life cycle of a medicine, and can be based on data from clinical trials or observational studies of the medicine, or safety findings from other medicines in the same class. In these cases, careful review of the available data can allow the regulator, often working with the developer, to develop a thoughtful and rational approach to drug safety issues in the postapproval period.

Reactive approaches are also needed in regulatory pharmacoepidemiology, because the adverse effects of medicine can become recognized at any time, sometimes many years, after approval. To the extent that regulators can use proactive pharmacoepidemiologic approaches, in theory reactive approaches can be minimized. However, not all drug safety issues can be predicted, so regulators will continue to need reactive approaches. These approaches require efficient review of the existing data, careful and timely assessment of the need for immediate or near-term regulatory action, and interaction with the product's manufacturer to plan further study. Reactive approaches become necessary, for example, when new safety issues are identified from spontaneously reported suspected adverse drug reactions, or when drug safety findings are published by independent groups, and neither the regulator nor the manufacturer is aware of them beforehand. Reactive approaches may also be needed when events such as manufacturing-related product recalls result in a large number of adverse event reports that need to be reviewed in a short period of time [21].

The specific scientific approach to an individual postapproval safety issue is beyond the scope of this chapter. From the regulator's point of view, the scientific studies that form the basis of regulatory actions must be as sound and robust as possible. Importantly, these studies should be designed to address the specific drug safety issue at hand, ideally in as short a time as possible. If there are ongoing planned studies, these may serve to contextualize the newly identified safety issue.

Assessing Actual Use Patterns of a Medicine

Regulators are interested not only in whether a medicine meets the relevant regulatory standards for approval, but also in how it is actually used in clinical practice. Because the harms of medicines can result not only from their intrinsic pharmacologic properties but also from how they are used, or misused, in practice, understanding the actual usage allows regulators to assess the degree to which the medicine is used in ways that are consistent with its safe use as described in the label or marketing authorization. To do so, regulators can use a variety of pharmacoepidemiologic techniques, including administrative claims data, electronic medical records, or other public health databases.

An analysis of emergency department visits in the US found that the rate of such visits involving both opioid analgesics and benzodiazepines increased from 11.0 to 32.4 per 100 000 population between 2004 and 2011. During that same period, overdose deaths involving drugs from both classes increased from 0.6 to 1.7 per 100 000 [22]. To shed light on this finding, an analysis of trends in the concomitant prescribing of opioids and benzodiazepines in the US between 2002 and 2014 found that concomitant prescribing increased by 41%, from 6.8% to 9.6% [23]. These drug utilization data were an important component of the body of evidence that led to warnings advising against the concomitant use of benzodiazepines and opioids [24].

An analysis of outpatient electronic medical records at a university hospital in Japan compared prescribed doses of agents for rheumatoid arthritis, diabetes, high blood pressure, and depression to standard approved doses [25]. The study found notable differences in the clinical characteristics of patients between the actual practice setting and the clinical trial setting. The average prescribed doses of agents for rheumatoid arthritis and depression were lower than standard approved doses, especially in older patients. The findings also suggested that the incidence of certain adverse events may differ between actual practice and clinical trials. Findings such as these, which describe how medicines are used in practice, can form the basis of more targeted drug safety studies.

Assessing the Impact of Regulatory Actions

Because of its public health focus, drug regulation must ensure that its actions lead to the intended public health outcomes. For serious safety issues, it is not enough simply to add a warning to a product label. Such an action is in itself an intervention, and it is thus important to understand its impact. Recognizing the fundamental importance of the need for such assessments, the Pharmacovigilance Risk Assessment Committee of the EMA developed a formal strategy to measure the impact of pharmacovigilance activities. The strategy is aimed both at informing the review of individual medicines that have been the subject of major risk minimization efforts and at determining which activities are successful and which are not, in order to optimize the pharmacovigilance system. Pharmacoepidemiology is critical to this endeavor, as it can relate regulatory activities to the outcomes that those activities are intended to affect. Pharmacoepidemiologic thinking and methodologies underpin the EMA's strategy [26].

One domain of assessment of the impact of regulatory activities is an understanding of the

effectiveness of regulatory agencies' communications about the risks of medicines. A study that examined the extent to which patients understand important information about a serious risk of a medicine that they are taking considered the results of patient-directed knowledge surveys for 66 medicines for which patients were supposed to have received a Medication Guide, a type of patient-directed labeling [27]. For each Medication Guide, acceptable knowledge was defined as 80% or more of patients correctly answering questions about the medicine's primary risk. The study found that only 20 Medication Guides (30.3%) met the 80% threshold, a finding that underscores the need for improved patient-directed information.

To understand the diffusion of an emerging drug safety message, a series of studies in the US employed several complementary methodologies, including quantitative and qualitative analysis of traditional and social media, patient and prescriber interviews, a patient survey, and pharmacoepidemiologic analysis of healthcare claims data [28]. These studies, which examined drug safety communications concerning the risk of next-day drowsiness and mental impairment with the sleep aid zolpidem and a recommendation to change the starting dose from 10 mg to 5 mg, found that traditional media reported widely some of the messages in the safety communication, but reported other messages less widely [29]. Semi-structured interviews of patients and physicians, which were designed to assess awareness and understanding of the messages, found that patients and physicians use a variety of sources of drug safety information, and that some of the messages in the drug safety information were communicated effectively to patients and physicians, though none of the patients had recalled hearing all of the messages and some patients did not understand fully how the message applied to them [30]. Studies such as these allow regulators to assess the impact of their communications and thus provide an opportunity for improvement of these efforts.

Analysis of the impact of regulatory actions is not limited to the assessment of actions related to individual medicines. Rather, it can look broadly at how the functioning of a drug regulatory system contributes to the system's public health mission. Because of the rapidly changing and expanding data that inform pharmacoepidemiologic studies, studies that examine overall performance are important because they can lead to system-wide improvements. For example, an analysis of 144 safety-related actions taken in 2012 by the Japanese Ministry of Health, Labour, and Welfare (MHLW) and the PMDA found that 83.5% were based on spontaneous reports [31]. The actions were not limited to recently approved medicines, since the median duration between drug approval and the safety-related regulatory action was 8 years. While the median duration between signal detection and a tentative or final decision was relatively short (49 and 84 days, respectively), assessments involving older products or multiple products required substantially more time. Studies such as this one can point to the need for the development of pharmacoepidemiologic assessment that utilizes health information databases in addition to spontaneous reports. A new program advanced by the PMDA, known as MIHARI (Medical Information for Risk Assessment Initiative), seeks to fulfill this need, and is described later in the chapter. In Japan, pharmacoepidemiologic studies using health information databases are permissible as studies for postapproval requirement purposes under the PMD Act. In addition, the MHLW and PMDA recently launched the Clinical Innovation Network (CIN) project, which aims to promote clinical studies by the use of disease registry data.

In another example, the Food and Drug Administration Amendments Act of 2007 required the US FDA to perform a summary safety analysis of newly approved medicines

18 months after approval or after 10 000 patients had taken the drug, whichever was later. These analyses were to be performed in addition to the FDA's ongoing safety monitoring of the drug. An evaluation of these summary safety analyses revealed that of 458 products that had been approved for at least 18 months, 300 had been the subject of a summary safety analysis; many products had not reached the 10 000-patient threshold and thus were not the subject of a summary safety analysis [32]. A new safety signal that resulted in a safety-related label change was found for 11 of these products. To provide context for this findings, these 11 safety-related label changes represented less than 2% of the 713 safety-related label changes made for these products. These findings suggested that the summary safety analysis had minimal value over standard pharmacovigilance activities. The FDA's requirement for these summary safety analyses was removed in a subsequent law [33].

Advancing the Science of Pharmacoepidemiology

Pharmacoepidemiology is a complex, dynamic, and changing field. It relies on the integration of epidemiology, clinical pharmacology, pharmacy, medicine, statistics, and other disciplines for its full execution. Increasingly, the rapid advances in the availability of large, diverse, and relevant datasets have made informatics an important contributing discipline to pharmacoepidemiologic efforts. Acquiring expertise in pharmacoepidemiology thus requires an environment that provides access to experts in all the relevant disciplines. Furthermore, this discipline relies on population-based healthcare data and thus an understanding of the healthcare system in which the data were generated, which experts in the above fields may not have. As more and more drug safety questions arise that require expertise in pharmacoepidemiology as well as appropriate data, it is crucial that there be sufficient capacity, both in the form of well-trained pharmacoepidemiologists and in the availability of systems, such as networks that combine relevant data with scientific expertise, that can be used for pharmacoepidemiologic studies. Because pharmacoepidemiology is a multidisciplinary effort, there must also be appropriate mechanisms for collaboration. Regulatory agencies play a role in facilitating the reaching of these goals.

To strengthen the monitoring of marketed medicines, the EMA developed the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) [34]. The EMA identified available expertise and research experience in the fields of pharmacovigilance and pharmacoepidemiology across Europe, and developed a network of centers with the capacity to perform postauthorization studies focusing on safety and benefit–risk. The ENCePP's principal activities include the development of a Code of Conduct to promote scientific independence and transparency [35]; the promulgation of scientific standards and guidance, including the ENCePP Guide on Methodological Standards in Pharmacoepidemiology [36], which provides guidance on methodology that is intended to supplement that in formal textbooks and regulatory guidance, and the ENCePP Checklist of Study Protocols [37], which is intended to promote the quality of pharmacoepidemiologic studies; and the ENCePP Resources Database, which provides an index of pharmacoepidemiologic research organizations, networks, and data sources in the EU [38]. The ENCePP project illustrates one way in which a regulatory agency can be involved in building pharmacoepidemiologic capacity.

The FDA, as one of its commitments under the reauthorization of the Prescription Drug User Fee Act in 2007, was tasked with developing a guidance document, with input from academia, industry, and others, "that addresses epidemiology best practices and provides guidance on carrying out scientifically sound

observational studies using quality data resources" [39]. The guidance document [40] illustrates another mechanism through which regulatory agencies can promote the field of pharmacoepidemiology.

Because pharmacoepidemiology depends on many areas of expertise, fostering collaboration to advance the field is another potential role that regulators can play. Through a project funded by the Innovative Medicines Initiative (IMI), the EMA, along with the national drug regulatory agencies from Denmark, Spain, and the UK, partnered with a number of public and private organizations, academic organizations, and pharmaceutical companies to form IMI PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium), a consortium dedicated to strengthening the methods used to monitor the benefits and risks of medicines [41]. (See also Chapter 25.) Topic areas covered by IMI PROTECT included enhancing data collection from consumers; improving early and proactive signal detection from spontaneous reports (including comparing different signal detection algorithms), electronic health records, and clinical trials; developing, testing, and disseminating methodologic standards for the design, conduct, and analysis of pharmacoepidemiologic studies; developing methods for continuous benefit–risk monitoring of medicines; and testing and validating various methods developed in PROTECT. By fostering collaboration across multiple disciplines, IMI PROTECT generated 74 original peer-reviewed scientific publications that not only advanced important pharmacoepidemiologic methodologies, but did so in a way that considered and enabled the potential regulatory application of such work [42].

Regulatory agencies can also promote pharmacoepidemiology by collaborating with external partners to develop and implement networks that allow for the regulatory agency to conduct its own specific pharmacoepidemiologic studies. In Japan, the PMDA launched the Medical Information for Risk Assessment Initiative (MIHARI), the aim of which is to use a variety of large-scale sources of electronic health information for pharmacoepidemiologic assessments [43]. *Mihari* is a Japanese word meaning "to monitor" or "to watch over." As part of its efforts to prepare the MIHARI for full-scale deployment, the PMDA has conducted more than 40 pilot studies during the past several years that made use of several large-scale databases: a nationwide insurance claims database covering nearly the entire Japanese population; a claims database maintained by health insurers covering a substantial portion of Japan; a database of electronic medical records maintained by medical institutions (e.g., the Medical Information Database NETwork, MID-NET®, described later in the chapter); and an inpatient care database containing data in a format compatible with the Diagnosis Procedure Combination (DPC) system, a comprehensive inpatient reimbursement system. Guided by the results of these pilot studies, in 2014 the PMDA introduced a novel regulatory framework for conducting postmarketing drug safety assessments using electronic health information databases. In addition to implementing this new framework, the PMDA also published the Guidelines for Pharmacoepidemiological Studies Using Health Information Databases for Drug Safety Assessments in 2014 [44]. The MID-NET system was established jointly by the MHLW and PMDA as a new database for use in drug safety assessment operations. It is a distributed database that compiles electronic medical records, insurance claims data, and DPC-compatible inpatient care data under a common data model. As of 2018, MID-NET had provided reliable data concerning approximately 4 million patients who received medical care from 23 medical institutions. In addition, in 2018, use of MID-NET was opened to relevant members of industry and academic researchers for use in pharmacoepidemiologic

studies investigating drug product safety and effectiveness.

FDA's Sentinel initiative (see also Chapter 25) also represents an example of a program sponsored by a regulatory agency that seeks to advance pharmacoepidemiology though a collaborative effort [45]. The goal of Sentinel is to create a sustainable, linked system of electronic healthcare databases to investigate safety questions about FDA-regulated medical products. The use of healthcare data in this way raises many questions of public interest, including on governance, privacy, data standards, and public disclosure of results. In view of these issues, the FDA sought extensive stakeholder input as it worked with outside organizations to develop Sentinel. In addition to the logistic issues already discussed, the fundamental premise of Sentinel – that data from many sources can be used to address a drug safety question in a timely way – implies that a collaborative effort is needed for the success of this project. The Sentinel system, the implementation of the initiative, includes administrative claims data and electronic medical record data from several holders of such data that are transformed into a common data model across all data holders. To facilitate efficient analysis of these data, the Sentinel system uses pretested, validated, and parameterized analytic programs along with the common data model. The system is designed as a distributed database in which holders of the data retain physical and operational control over their electronic data, a feature that addresses privacy concerns [46]. The Sentinel system has been used for a variety of analyses involving drugs as well vaccines.

Pharmacoepidemiologic efforts such as ENCePP, MIHARI, and Sentinel all make use of various traditional sources of healthcare data derived from existing sources that reflect current clinical practice and actual patient experiences. While these contemporary systems often rely on large datasets and, at times, integration of datasets through networks, it is important to note that pharmacoepidemiologic research has a decades-long tradition of using observational data recorded at the point of care to describe the effects of medicines in populations, though the scope of this work has largely focused on safety issues. The emergence in recent years of additional digital health-related data, such as data generated from wearable devices and health-related applications, has given rise to the notion that the expanding variety of electronic healthcare data can be used to study the effects of medicines beyond those related to safety. To promote further progress in this emerging area, numerous regulatory agencies from around the world resolved to work together to convert real-world data into real-world evidence to support regulatory decision making at the 12th Summit of Heads of Medicines Regulatory Agencies held in October 2017 in Kyoto, Japan. As methods to develop real-world evidence move forward [47], pharmacoepidemiologists, including those in regulatory agencies, will have an important and growing role to play, particularly in supporting timely and robust public health decisions.

References

1 Hunter NL, Rao GR, Sherman RE. Flexibility in the FDA approach to orphan drug development. *Nat Rev Drug Discov* 2017; **16**: 737–38. doi: 10.1038/nrd.2017.151

2 Braun MM, Farag-El-Massah S, Xu K, Cote TR. Emergence of orphan drugs in the United States: a quantitative assessment of the first 25 years. *Nat Rev Drug Discov* 2010; **9**: 519–22. doi: 10.1038/nrd3160

3 Ministry of Health, Labour and Welfare. Overview of orphan drug/medical device designation system. 2009. http://www.mhlw.

go.jp/english/policy/health-medical/pharmaceuticals/orphan_drug.html (accessed 19 December 2017)

4 European Medicines Agency. Orphan designation. 2015. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000029.jsp&murl=menus/regulations/regulations.jsp&mid=WC0b01ac05800240ce&jsenabled=true (accessed 9 December 2017).

5 Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. *Lancet* 2008; **371**: 2039–41. doi: 10.1016/S0140-6736(08)60872-7

6 Kaplan W, Wirtz VJ, Mantel-Teeuwisse A, Stolk P, Duthey B, Laing R. *Priority medicines for Europe and the world: 2013 update*. Geneva: World Health Organization, 2013. http://www.who.int/medicines/areas/priority_medicines/MasterDocJune28_FINAL_Web.pdf (accessed 19 December 2017).

7 Sherman RE, Li J, Shapley S, Robb M, Woodcock J. Expediting drug development – the FDA's new "breakthrough therapy" designation. *N Engl J Med* 2013; **369**: 1877–80. doi: 10.1056/NEJMp1311439

8 US Food and Drug Administration. Guidance for industry: expedited programs for serious conditions – drugs and biologics. 2017. https://www.fda.gov/media/86377/download (accessed 22 July 2019).

9 European Medicines Agency. PRIME: priority medicines. 2016. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000660.jsp (accessed 19 December 2017).

10 Fujiwara Y. Evolution of frameworks for expediting access to new drugs in Japan. *Nat Rev Drug Discov* 2016; **15**: 293–4. doi: 10.1038/nrd.2016.68

11 Temple RJ. A regulatory authority's opinion about surrogate endpoints. In Nimmo WS, Tucker GT, eds. *Clinical Measurement in Drug Evaluation*. New York: John Wiley, 1995.

12 US Food and Drug Administration. FDA facts: biomarkers and surrogate endpoints. https://www.fda.gov/about-fda/innovation-fda/fda-facts-biomarkers-and-surrogate-endpoints (accessed May 2019).

13 US Code of Federal Regulations. 21 CFR 314.510.

14 US Food and Drug Administration. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. 2018. https://www.fda.gov/media/71195/download (accessed 22 July 2019).

15 European Medicines Agency. Conditional marketing authorisation: report on ten years of experience at the European Medicines Agency. 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2017/01/WC500219991.pdf (accessed 19 December 2017).

16 Jokura Y, Yano K, Yamato M. Comparison of the new Japanese legislation for expedited approval of regenerative medicine products with the existing systems in the USA and European Union. *J Tissue Eng Regen Med* 2018; **12**(2): e1056–e1062. doi: 10.1002/term.2428

17 Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 2012; **10**: 2973–84. doi: 10.1002/sim.5403

18 US Food and Drug Administration. Abuse-deterrent opioids – evaluation and labeling: guidance for industry. 2015. https://www.fda.gov/media/84819/download (accessed 22 July 2019).

19 US Department of Health and Human Services, Food and Drug Administration. Guidance for industry: diabetes mellitus – evaluating cardiovascular risk in new antidiabetic therapies to treat Type 2 diabetes. 2018. https://www.fda.gov/media/71297/download (accessed 22 July 2019).

20 International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH

harmonized tripartite guideline: pharmacovigilance planning (E2E). 2004. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2E/Step4/E2E_Guideline.pdf (accessed 19 December 2017)

21 McMahon AW, Pratt RG, Hammad TA, *et al*. Description of hypersensitivity adverse events following administration of heparin that was potentially contaminated with oversulfated chondroitin sulfate in early 2008. *Pharmacoepidemiol Drug Saf* 2010; **19**(9): 921–33. doi: 10.1002/pds.1991

22 Jones CM, McAninch JK. Emergency department visits and overdose deaths from combined use of opioids and benzodiazepines. *Am J Prev Med* 2015; **49**: 493–501. doi: 10.1016/j.amepre.2015.03.040

23 Hwang CS, Kang EM, Kornegay CJ, Staffa JA, Jones CM, McAninch JK. Trends in the concomitant prescribing of opioids and benzodiazepines, 2002–2014. *Am J Prev Med* 2016; **51**: 151–60. doi: 10.1016/j.amepre.2016.02.014

24 US Food and Drug Administration. FDA Drug Safety Communication: FDA warns about serious risks and death when combining opioid pain or cough medicines with benzodiazepines; requires its strongest warning. 2016. https://www.fda.gov/Drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-about-serious-risks-and-death-when-combining-opioid-pain-or (accessed 22 July 2019).

25 Tanemura N, Uyama Y, Nagashima K, *et al*. Comparison of drug use between clinical practice and regulatory approval: results in older Japanese patients with rheumatoid arthritis, diabetes, high blood pressure, or depression. *Ther Innov Regul Sci* 2016; **50**: 743–50. doi: 10.1177/2168479016648731

26 Goedecke T, Morales D, Pacarariu A, Kurz X. Measuring the impact of medicines regulatory interventions – systematic review and methodological considerations. *Br J Clin Pharmacol* 2017; **84**(3). doi:10.1111/bcp.13469

27 Knox C, Hampp C, Willy M, Winterstein AG, Dal Pan G. Patient understanding of drug risks: an evaluation of medication guide assessments. *Pharmacoepidemiol Drug Saf* 2015; **24**: 518–25. doi: 10.1002/pds.3762

28 Kesselheim AS, Campbell EG, Schneeweiss S, *et al*. Methodological approaches to evaluate the impact of FDA drug safety communications. *Drug Saf* 2015; **38**: 565–75. doi: 10.1007/s40264-015-0291-y

29 Woloshin S, Schwartz LM, Dejene S, *et al*. Media coverage of FDA drug safety communication about zolpidem: a quantitative and qualitative analysis. *J Health Commun* 2017; **22**: 365–72. doi: 10.1080/10810730.2016.1266717

30 Kesselheim AS, McGraw SA, Dejene SZ, *et al*. Patient and physicians perceptions of drug safety information for sleep aids: a qualitative study. *Drug Saf* 2017; **40**: 531–42. doi: 10.1007/s40264-017-0516-3

31 Ishiguro C, Misu T, Iwasa E, Izawa T. Analysis of safety-related regulatory actions by Japan's pharmaceutical regulatory agency. *Pharmacoepidemiol Drug Saf* 2017; **26**: 1314–20. doi: 10.1002/pds.4252

32 Sekine S, Pinnow EE, Wu E, Kurtzig R, Hall M, Dal Pan GJ. Assessment of the impact of scheduled postmarketing summary safety analyses on regulatory actions. *Clin Pharmacol Ther* 2016; **100**: 102–08. doi: 10.1002/cpt.346

33 21st Century Cures Act, Public Law 114-255, December 13, 2016.

34 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. About ENCePP. www.encepp.eu (accessed 19 December 2017).

35 European Medicines Agency. The ENCePP Code of Conduct. 2016. http://www.encepp.eu/code_of_conduct/documents/ENCePPCodeofConduct.pdf (accessed 19 December 2017).

36 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology (revision 6). EMA/95098/2010. 2010. http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml (accessed 19 December 2017).

37 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). ENCePP checklist for study protocols (revision 3). EMA/540136/2009. 2009. http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml (accessed 19 December 2017).

38 European Network for Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). ENCePP resources database. http://www.encepp.eu/encepp/resourcesDatabase.jsp (accessed 19 December 2017).

39 Department of Health and Human Services. US Food and Drug Administration. FDA PDUFA IV drug safety five-year plan, December 2008. https://www.fda.gov/media/74936/download (accessed May 2019).

40 US Food and Drug Administration. Guidance for industry and FDA staff: best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. 2013. https://www.fda.gov/media/79922/download (accessed 22 July 2019).

41 PROTECT. Welcome to the PROTECT website! http://www.imi-protect.eu (accessed 19 December 2017).

42 European Medicines Agency. Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT): results and their impact on regulatory practice. 2016. http://www.imi-protect.eu/documents/ReportImpactPROTECT_002.pdf (accessed 19 December 2017).

43 Ishiguro C, Takeuchi Y, Uyama Y, Tawargi T. The MIHARI project: establishing a new framework for pharmacoepidemiological drug safety assessments by the Pharmaceuticals and Medical Devices Agency of Japan. *Pharmacoepidemiol Drug Saf* 2016; **25**: 854–9. doi: 10.1002/pds.4032

44 Pharmaceuticals and Medical Devices Agency. Guidelines for pharmacoepidemiological studies using health information databases for drug safety assessments. 2014. https://www.pmda.go.jp/files/000147250.pdf (accessed 19 December 2017). (Japanese)

45 US Department of Health and Human Services, US Food and Drug Administration. The Sentinel Initiative – national strategy for monitoring medical product safety. 2008. http://www.fda.gov/media/75240/download (accessed 22 July 2019).

46 Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's Sentinel Initiative – a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther* 2016; **99**: 265–8. doi: 10.1002/cpt.320

47 Sherman RE, Anderson SA, Dal Pan GJ, *et al*. Real-world evidence – what is it and what can it tell us? *N Engl J Med* 2016; **375**: 2293–7. doi: 10.1056/NEJMsb1609216

9

Pharmacoepidemiology and the Law

Aaron S. Kesselheim and Kerstin N. Vokinger

Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

The law describes the basic rules under which people live in modern society. Tort law, for example, provides a system of corrective justice and a coherent set of principles to decide whether a person deserves compensation for an injury he or she sustained. As another example, contract law provides a structure for adjudicating agreements between parties. In both cases, the existence of governing law helps influence the way people act. In the case of tort law, knowledge about its liability rules should incentivize people to take care to prevent accidents from happening.

In their daily work, pharmacoepidemiologists encounter many different aspects of the law. Perhaps the most recognizable connection occurs when patients seek redress in tort law for adverse effects from a medical product. In such circumstances, pharmacoepidemiologic studies may provide the scientific underpinning for the claim as to the association between the drug and the claimed outcome. Often, pharmacoepidemiologists are called as expert witnesses to interpret scientific findings for judges and juries. Other basic legal principles may also have important effects on the practice of pharmacoepidemiology. For example, pharmacoepidemiologists must navigate contract law when they develop research agreements with funding sources or owners of databases. Pharmacoepidemiologists interface with property law when they attempt to secure ownership rights over their discoveries using patents (a type of "intellectual property").

This chapter outlines three of the most recognizable intersections of pharmacoepidemiology and the law: tort law, contract law, and intellectual property law. The chapter defines and describes basic legal rules in these subject areas, and uses these rules as a basis for additional discussion about practical and ethical implications for pharmacoepidemiology. In each example, US law is used as the paradigm, with some attention to alternative models in Europe. Since much of the discussion is based on principles that are generally similar in other comparable legal systems, the lessons are applicable to pharmacoepidemiologists around the world.

Tort Law and Product Liability Lawsuits

Product liability lawsuits provide an opportunity for individuals harmed by a drug to seek damages from its manufacturer. Recent widely

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

reported cases have included the nonsteroidal anti-inflammatory drug (NSAID) rofecoxib (Vioxx®), the antidepressant paroxetine (Paxil®) and other selective serotonin reuptake inhibitors (SSRIs), olanzapine (Zyprexa®) and other atypical antipsychotics, the cholesterol-lowering agent cerivastatin (Baycol®), the antidiabetic/anti-inflammatory troglitazone (Rezulin®), and the serotoninergic anorectic drug dexfenfluramine (Redux®). In this chapter, we will review how product liability lawsuits are adjudicated according to some common law principles. A basic understanding of product liability law is essential for pharmacoepidemiologists, even for those who might never find themselves in a courtroom, because such lawsuits also exert substantial influence on the field. Tort litigation brought by government agencies and individual patients can help uncover previously unavailable data on adverse effects, questionable practices by manufacturers, and flaws in drug regulatory systems [1].

The Legal Theory of Product Liability

In the centuries-old common law tradition of England, which forms the basis for legal systems in the US and a number of other countries, a consumer injured by a defective or contaminated pharmaceutical product was not permitted a right of action unless the consumer purchased the preparation directly from the manufacturer. The emergence of product liability law altered that state of affairs, permitting consumers harmed by the many products sold widely in interstate commerce and through distributors like pharmacies to seek redress for their injuries from the original manufacturers [2]. Originally, product liability was grounded in the theory of negligence, which meant that defendants would be liable for causing plaintiffs' injuries if the defendants engaged in wrongful or unreasonable conduct, even if it was unintentional. To succeed in a claim for negligence, plaintiffs needed to show (i) that defendants had a duty to exercise reasonable care; (ii) that defendants' conduct diverged from customary practices that would be followed by other manufacturers or members of the industry; (iii) that there was a causal link between the defendants' lack of care and the outcome at issue; and (iv) that the preceding three factors led to damages.

However, negligence theory did not allow enough deserving plaintiffs to be compensated for product-related injuries they suffered, particularly in cases in which products were hazardous or dangerous. Judges rationalized that some products contained an inherent risk of harm, so manufacturers that chose to sell such products needed to bear the responsibility when the products caused injury. As a result, starting in the early 1960s, judges started applying the theory of strict liability to certain product liability cases. Strict liability merely requires demonstration that the dangerous product caused the injury; as distinguished from negligence, the question is moot as to whether the defendants followed customary practices or exercised reasonable precautions. This principle permitted plaintiffs to seek compensation for injuries merely because the product was designed a certain way, irrespective of other mitigating factors. For example, the product could have a "manufacturing defect," meaning that the product did not comply with the manufacturer's own standards, or a "design defect," meaning that the product was designed in a way that conferred inherently unreasonable risk for the consumer.

Strict product liability grew quickly in popularity. In 1965, US legal scholars proposed a consensus understanding of the area in the influential Restatement (Second) of Torts, finding that a seller of a product that is "in a defective condition unreasonably dangerous to the user or consumer" should be strictly liable even if the seller "exercised all possible care in the preparation and sale of the product" [3]. Notably, the authors commented that warnings could be employed to prevent any product from being deemed "unreasonably dangerous," although

such warnings needed to address risks that the seller "has knowledge, or by application of reasonable, developed human skill and foresight should have knowledge" [4]. Thus, strict product liability also allowed plaintiffs to bring causes of action against manufacturers based on inadequate warnings, otherwise known as a "failure to warn."

Some courts were hesitant to apply strict product liability to cases emerging from the pharmaceutical field. This reticence was reflected in the Restatement, which included an important annotation relevant to prescription drugs. In Comment k to this section of the document, the Restatement noted that a pharmaceutical product "properly prepared, and accompanied by proper directions and warning, is not defective, nor is it unreasonably dangerous" [5]. Thus, the Restatement excluded most prescription drugs from strict liability based on manufacturer or design defects. The authors separated pharmaceutical products from other products because they believed the marketing and use of pharmaceutical products "are fully justified, notwithstanding the unavoidable high degree of risk which they involve." Prominent legal scholar William Prosser summed up the justification for treating prescription drugs differently:

> The argument that industries producing potentially dangerous products should make good the harm, distribute it by liability insurance, and add the cost to the price of the product, encounters reason for pause, when we consider that two of the greatest medical boons to the human race, penicillin and cortisone, both have their dangerous side effects, and that drug companies might well have been deterred from producing and selling them. Thus far the courts have tended to hold the manufacturer to a high standard of care in preparing and testing drugs of unknown potentiality and in giving warning; but in the absence of evidence that this standard has not been met, they have refused to hold the maker liable for unforeseeable harm. [6]

Ultimately, a minority of US courts have implemented the Comment k principle and offered pharmaceutical manufacturers a blanket protection from strict liability for manufacturer or design defect claims [7]. The majority of courts charted a slightly different course. For example, in New Jersey, the state Supreme Court declined to adopt Comment k in the case of an infant who suffered severe tooth discoloration after being prescribed demeclocycline (Declomycin®), a tetracycline antibiotic. The court ruled that the Comment k shield should only apply to drugs that were "more vital to the public health and human survival than others," while less useful drugs would continue to be evaluated under strict liability [8].

In 1997, the Restatement (Third) of Torts: Product Liability tried to clarify the question about liability for design defects. It reemphasized that judicial risk–utility analysis was improper, arguing that a drug cannot be considered to have a design defect if "reasonable health care providers, knowing of such foreseeable risks and therapeutic benefits" prescribed the drug to the patient [9].

Even in jurisdictions amenable to strict product liability for pharmaceuticals, the vast majority of drugs approved by the US Food and Drug Administration (FDA) are likely to meet courts' balancing test. As a result, when a person is injured by a prescription drug, a "design defect" lawsuit based on the claim that the product was avoidably unsafe is very unlikely to succeed. Rather, plaintiffs usually seek to demonstrate "failure to warn" by the manufacturer about the adverse event at issue (nominally a strict liability claim). Alternatively, plaintiffs could sue based on a negligence theory that the manufacturer failed to take reasonable care in marketing its product, an analysis that also largely hinges on the appropriateness of the accompanying warnings. Practically speaking, the ultimate

disposition of a case filed under a strict liability failure to warn or negligence theory turns on the question of whether the warning is reasonable [10]. After these historical twists and turns in legal theory in this area, the claim for "failure to warn" has become the most common basis for litigation over pharmaceutical products. In the next section, we will review the link between the work of pharmacoepidemiologists and failure-to-warn claims.

Failure-to-Warn Claims

Whether based on strict liability or negligence, a failure-to-warn product liability action includes three main contentions: (i) knowledge of the drug risk by the manufacturer; (ii) improper warning of the drug risk; and (iii) causation of damages.

Knowledge of the Drug Risk by the Manufacturer

First, the plaintiff must demonstrate that a pharmaceutical manufacturer knew, or should have known, of the risk. Apart from the rare case decided based on a strict liability design defect, a manufacturer of a pharmaceutical product is not held accountable for risks about which it could not have known. For example, in one case, a plaintiff brought a lawsuit claiming that her oral contraceptive medication led to her having a cerebrovascular accident, or stroke [11]. The court remarked, "Dates are thus vitally important as there is no duty to warn of unknown or unforeseeable risks, and the question is whether the risk was knowable or reasonably foreseeable at the time when the plaintiff was still taking the drug." The jury found that the particular risk the plaintiff claimed could not have been known at the time the drug was prescribed, based in part on the testimony of the expert pharmacoepidemiologist who reported that "new techniques to measure these clotting effects had not then been developed" at the time of the injury. According to the court, "The warnings contained in the package inserts were adequate or … the statements contained therein were a fair representation of the medical and scientific knowledge available at the time the drug was taken by the plaintiff."

Knowledge can be actual or constructive. *Actual knowledge* is defined as literal awareness. Actual knowledge can be demonstrated by showing that the manufacturer was cognizant of reasonable information suggesting a particular risk that it did not pass on to consumers, for example when a defendant possesses data about relevant adverse events that were not disclosed. In the case of SSRIs used to treat depression, various manufacturers were found to have conducted clinical trials that showed an increased risk of suicidal ideation in adolescent patients taking the drug. Plaintiffs brought lawsuits charging that these findings were knowingly delayed for lengthy periods of time, not released, or the concerns not fairly represented [12]. For example, the largest study of paroxetine (Paxil®) in pediatric patients was conducted in the US from 1993 to 1996; it showed no benefit of the drug over placebo and 5 cases (out of 93) of suicidal ideation, as compared to 1 case out of 89 in the placebo arm and 1 case out of 95 in the comparator (non-SSRI) arm. The manufacturer, GlaxoSmithKline, allegedly sought to "effectively manage the dissemination of these data in order to minimize any potential negative commercial impact" [13]. To support this contention, plaintiffs pointed to the fact that the data were only presented in abstract form in 1998 and published in 2001 (when the authors concluded that the drug was "generally well tolerated and effective for major depression in adolescents") [14]. After the full data from this trial and others like it were made public, a new FDA health advisory in 2004 warned physicians to carefully monitor patients for "clinical worsening, as well as agitation, irritability, suicidality, and unusual changes in behavior" and emphasized that only the SSRI fluoxetine

(Prozac®) had been approved to treat pediatric major depressive disorder [15].

Constructive knowledge is sometimes called "legal knowledge," because it is knowledge that the law assumes should be present, even if it is not. Constructive knowledge is knowledge that a person did not have, but could have acquired by the exercise of reasonable care. For example, the cholesterol-lowering drug cerivastatin (Baycol®) was removed from the market in 2001 after it was linked to cases of rhabdomyolysis, a potentially fatal kidney disease. The manufacturer, Bayer, was found to possess several reports from as early as 1999 suggesting a 10-fold risk of rhabdomyolysis relative to other medications in its class, but it allegedly did not process these reports and pass them along to patients or regulators [16]. A memorandum from a Bayer official stated, "If the FDA asks for bad news, we have to give [it], but if we don't have it, we can't give it to them" [17]. In this case, Bayer could be said to have constructive knowledge of these concerns by 1999, because the company should have processed the reports and acted on them. In other cases, plaintiffs have tried to prove constructive knowledge by arguing that manufacturers should have performed different or additional analyses to better understand an important side effect of their product. The standard for constructive knowledge in these situations has been what a reasonably prudent company with expertise in this area would have undertaken.

Improper Warning of the Drug Risk

If a manufacturer has the duty to provide a warning about adverse events associated with its product, then the next question is whether an adequate warning was provided. A proper warning has certain hallmarks, including relevance, timeliness, and accuracy.

First, a warning about an adverse effect must be commensurate with the scope and extent of dangers associated with the drug. In the case of troglitazone (Rezulin®), an oral hypoglycemic approved in the US in 1997 and used by diabetic patients, the company was accused of minimizing its presentation of liver toxicity in its warning materials [18]. Elevations of hepatic enzymes in early testing were initially depicted in the descriptions of adverse effects simply as "≥3-fold." Yet, some were apparently more than 20-fold; several of those patients suffered acute liver failure. In the subsequent litigation, it was alleged that the warning was deficient because company did not initially acknowledge this clinically important difference [19].

Secondly, warnings must not be subject to undue delay. Some delays may be internal. In the case of rosiglitazone (Avandia®), another oral hypoglycemic drug, a 2007 meta-analysis linked the drug to life-threatening cardiovascular adverse events [20]. However, after a review of internal company documents, a US Senate Finance Committee report suggested that the manufacturer knew about these risks years before this article was published, but delayed warning about them and sought to limit their dissemination [21]. A primary question in lawsuits arising from the use of rosiglitazone is whether these tactics inappropriately delayed reasonable warnings about the adverse effect. Sometimes, interactions with regulators may cause delays. For example, cisapride (Propulsid®) was a pro-kinetic agent linked to potentially fatal cardiac side effects. It was reported that the manufacturer and the FDA negotiated for five years over the details of how to change the drug's label to include adverse event data that had been submitted to the agency but not made fully available to the public [22].

Thirdly, warnings must be of appropriately urgent tone. In the case of rofecoxib (Vioxx®), a new type of NSAID used for arthritis, preapproval clinical trials suggested enhanced risk of serious cardiovascular side effects, a result consistent with a later pivotal manufacturer-sponsored trial comparing the drug to naproxen, another older NSAID, in a population of patients with rheumatoid arthritis (but no

known cardiovascular disease) [23]. When the drug's official FDA label was updated in 2002 to account for these findings, subsequent lawsuits alleged that the warning was insufficiently urgent because the risk of cardiovascular events was described in vague terms and placed in the less prominent "precautions" section of the labeling [24].

Finally, a manufacturer's duty does not end with the initial warning, because it must keep up with emerging scientific data and patient reports, and warn of new side effects discovered after initial approval. In one case, plaintiffs brought a suit contending that their daughter's serious birth defects were related to a teratogenic progesterone formulation (Delalutin®) manufactured by the defendant. The court noted that the drug manufacturer is under a "continuous duty … to keep abreast of scientific developments touching upon the manufacturer's product and to notify the medical profession of any additional side effects discovered from its use" [25]. The plaintiff's expert medical witness testified that there was "sufficient scientific information and literature relative to progesterones" at the time the drug was used to "make a prudent drug manufacturer do teratogenicity studies on any progesterone agent" [25].

Causation of Damages

Another major issue in a pharmaceutical product liability case is whether the product at issue actually caused the alleged injury. Pharmacoepidemiologists may be most comfortable thinking about causation from a medical or scientific point of view. Scientists generally posit hypotheses to explain particular outcomes and then test those hypotheses by studying whether variations in the outcomes exist across populations. However, legal causation usually requires a clear causal chain from exposure to outcome, in an individual. The legal standard for causation is therefore challenged by product liability cases, in which probabilistic evidence (i.e., P values or confidence intervals) often links drugs to injuries [26]. Courts must address two types of legal causation: general and specific causation.

General causation addresses whether a product *can* cause a particular injury in the population of patients like the plaintiff. The common law standard to prove general causation is that a particular product "more likely than not" caused the damages. Some courts have held that legal causation must be demonstrated by more than an association and a mere possibility of causation, even though causal hypotheses based on such considerations are common in the scientific literature. A few courts have even gone further and defined "more likely than not" as having a relative risk of greater than 2.0, no matter how tight the confidence intervals are around a statistically significant finding of association between 1.0 and 2.0 [27]. Presumably this is based on the calculation of attributable risk in the exposed group exceeding 50%, when the relative risk exceeds 2.0. This standard has been replicated in the Federal Judicial Center's *Reference Manual on Scientific Evidence* [28] and employed in some cases to exclude epidemiologic evidence with weaker associations. For example, in the case of the antinausea drug pyridoxine/doxylamine (Bendectin®), which was claimed to be causally linked with birth defects, one court noted, "In terms of statistical proof … plaintiffs must establish not just that their mothers' ingestion of Bendectin increased somewhat the likelihood of birth defects, but that it more than doubled it – only then can it be said that Bendectin is more likely than not the source of their injury" [29]. In one case related to litigation over the link between silicone breast implants and inflammatory disease, a court excluded a study linking the product and the outcome with a relative risk of 1.24, noting that the finding was "so significantly close to 1.0" that the study "was not worth serious consideration for proving causation" [30].

However, all courts do not adhere rigidly to the 2.0 relative risk principle for general causation.

Both clinical trials and epidemiologic studies of the product at issue can establish general causation between a pharmaceutical product and an outcome. Animal studies, meta-analyses, case reports/case series, and secondary source materials (such as internal company documents) have been used in court as they are in the medical field – to help support establishing a causal link. Since pharmacoepidemiologic studies tend to assess the presence of an association, rather than directly addressing causation, courts sometimes apply the Bradford Hill criteria to connect an association with general causation (see Table 9.1 and Chapter 1).

To demonstrate *specific causation*, a plaintiff must show that the product in question caused the alleged injury in the plaintiff. This can be a particularly complex issue for pharmaceutical products. In some cases, like instantaneous allergic reactions, the causal link between a product and an outcome is clear. For more subacute or later-onset responses, however, specific causation may be hard to demonstrate. For example, in one case against Merck brought by a plaintiff who suffered a myocardial infarction shortly after starting rofecoxib, the manufacturer argued that the outcome was attributable to the plaintiff's prior existing coronary artery disease. The plaintiff countered with the fact that he was in a state of stable cardiovascular health prior to initiation of rofecoxib, that he simultaneously developed two coronary artery clots after the drug's initiation (a rare presentation for ischemic heart disease), and that many studies have confirmed the link between rofecoxib and cardiovascular disease (a point relevant to general causation) [31]. While the trial court held for the plaintiff, the decision was reversed on appeal; the appeals court ruled that, "although plaintiffs were not required to establish specific causation in terms of medical certainty, nor to conclusively exclude every other reasonable hypothesis, because [the plaintiff's] preexisting cardiovascular disease was another plausible cause of his death, the plaintiffs were required to offer evidence excluding that cause with *reasonable certainty*" [32].

Another important aspect of specific causation is that the plaintiff must demonstrate that the inadequate warnings about the adverse effect were relevant to the plaintiff's receiving the drug. If a defendant can demonstrate that even an adequate warning would have made no difference in the decision to prescribe the drug, or to monitor the patient postprescription, the case may be dismissed for lack of a proximate cause.

Table 9.1 Bradford Hill criteria.

1)	Strength of association
2)	Consistency and replication of findings
3)	Specificity with respect to both the substance and injury at issue
4)	Temporal relationship
5)	Biological gradient and evidence of a dose–response relationship
6)	Plausibility
7)	Coherence
8)	Experimental removal of exposure
9)	Consideration of alternative explanation

Source: Adapted from Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**: 295–300. Reproduced with permission of SAGE Publications.

Learned Intermediary Defense

If a plaintiff successfully argues these issues and demonstrates a *prima facie* case of product liability based on a failure to warn, the manufacturer has a few possible defenses. The most relevant in the field of pharmaceutical law is the learned intermediary defense.

Originally, product liability law imposed on all manufacturers a duty to warn consumers about the risks of their products. However, starting in the 1960s, pharmaceutical manufacturers argued that it would be more effective for them to warn physicians, the gatekeepers of prescription

medicines [33]. Courts accepted that physicians' advanced training and direct contact with patients put them in an optimal position to understand and relay complex information about possible side effects. Physicians are also well placed to discuss risks and benefits applicable to particular clinical circumstances in their patients. The "learned intermediary" rule allows pharmaceutical manufacturers to fulfill their duty to warn by providing an accurate and adequate warning to prescribing physicians [34].

The implications of the learned intermediary defense are that the debates in plaintiffs' cases tend to focus on the propriety of the warning vis-à-vis the physician, rather than the patient. Therefore, warnings do not have to be offered about risks that should be obvious or are generally known to skilled medical practitioners [35]. However, when the information given to physicians omits, underemphasizes, misstates, or obfuscates dangers, this deficiency is legally transferred to the patient, who maintains a right of redress against the manufacturer if those dangers materialize and cause injury.

If the manufacturer imparts an appropriate warning to physicians, then the manufacturer can be insulated from liability. In such cases, the focus of the litigation then often turns to the conduct of the physician and the physician–patient interaction. For example, in one case a lawsuit was brought following the suicide of a patient who had been prescribed two antihypertensive drugs, hydrochlorothiazide (HCTZ) and reserpine (Harmonyl®). The label for HCTZ stated that it might "potentiate the action of other antihypertensive drugs," while the insert for reserpine stated that the drug should be discontinued at any sign of "despondence" and that there were reports of drug-related depression severe enough to result in suicide. Because the physician was presumed to have had constructive knowledge of both of these warnings, the court insulated the manufacturers from liability [36].

In special situations, pharmaceutical manufacturers may lose the ability to invoke the learned intermediary defense. If a manufacturer markets its product very aggressively and without sufficient attention to certain risks, courts may rule that it has essentially undone the physician–patient prescribing relationship. Direct-to-consumer advertising (DTCA) is one modality that can undercut the assumption that patients are largely ignorant of prescription drug risks and that manufacturers lack means of interacting with patients other than through physicians. DTCA is currently only permitted in two industrialized countries around the world: the US and New Zealand. The New Jersey Supreme Court has ruled that DTCA created a limited exception to the learned intermediary defense [37], and in 2007 the West Virginia Supreme Court rejected the learned intermediary defense in its entirety on this basis [38]. Nonetheless, in most jurisdictions, the learned intermediary rule still stands.

Expertise and *Daubert*

Pharmacoepidemiologists often serve as expert witnesses in product liability cases. Pharmacoepidemiologists can help judges and juries understand data about drugs and help determine whether warning information appropriately reflects the risk posed by a drug. Experts are usually called on to describe the current state of knowledge about the adverse event at issue, and may be asked to perform additional pharmacoepidemiologic analyses of available data to present before the court.

However, courts can exclude some practitioners and some analyses from trial. Traditionally, the judge is responsible for evaluating whether expert witnesses lack qualifications or espouse scientific theories out of step with accepted knowledge [39]. In the 1993 case of *Daubert v. Merrell Dow*, the US Supreme Court outlined a number of criteria for reviewing the appropriateness of expert witness testimony, including whether the theory was current and whether it had been tested or subjected to peer review and publication [40]. A subsequent case applied these

rules and further refined them in evaluating a debate over the admissibility of expert testimony suggesting that polychlorinated biphenyls (PCBs) can cause lung cancer. The research was excluded because the experts did not validate their conclusions: the epidemiologic studies did not report a statistically significant causal link between PCBs and lung cancer, lacked proper controls, and examined substances other than PCBs [41]. As federal circuit court judge Richard Posner has explained in separate circumstances, "the courtroom is not the place for scientific guesswork, even of the inspired sort" [42].

In the US, some state courts have embraced the *Daubert* guidelines, which have also been taken up by revised Federal Rules of Evidence [43]; others adhere to an alternative doctrine that excludes testimony containing theories that do not enjoy "general acceptance in the relevant scientific community" [44]. Thus, pharmacoepidemiologists seeking to present expert evidence in litigation will routinely face judicial inquiry to determine whether they are fit to serve in that role. Judicial oversight in general sets a low floor for reliable expert testimony, although it can be expected to exclude experts who lack the relevant qualifications, lack facts to back up their perspectives, lack reliable methods, or fail to apply the methods appropriately [45]. There is considerable skepticism about the effectiveness of courts as a gatekeeper for expert witnesses, with some commentators citing judges' lack of the technical knowledge needed to meaningfully evaluate medical and scientific expertise [46].

The Effect of Regulation on Product Liability Litigation in the US

In the last few years, there has been a wave of controversy about the role of government regulation of pharmaceuticals in product liability claims against drug manufacturers. Under the US Food, Drug, and Cosmetic Act, originally passed in 1938, the FDA is required to certify that prescription drugs are safe enough and show efficacy for their intended indication before being sold on the US market [47] (see also Chapters 1 and 8). At the time of approval, the FDA also endorses the official drug labeling, which presents a description of the basis for the drug's efficacy as well as safety concerns that have emerged during the preapproval testing [48]. The labeling, which is generally written by the manufacturer and approved by the FDA, has legal significance as well. For example, because the FDA restricts certain types of manufacturer communication about non-FDA-approved (or "off-label") indications, the label determines what a pharmaceutical manufacturer can communicate to physicians and the public about its product [49]. The FDA requires the manufacturer to mention important warnings that are in the official labeling when marketing its product, but does not require manufacturers to mention warnings that are not in the labeling.

For most of its history, the FDA has regulated the drugs sold in the US without any direct role in product liability litigation brought by consumers injured by FDA-approved drugs [50]. The agency's noninterventionist posture changed for the first time in September 2002 in a product liability case brought after a man was prescribed the SSRI sertraline (Zoloft®) and started experiencing agitation, confusion, and suicidal thinking, ultimately leading him to take his own life one week later [51]. The plaintiffs claimed that the manufacturer failed to warn appropriately about the risks of suicide. The manufacturer contended that such a claim could not be brought because the FDA had not included such a warning in the official label, and the Supremacy Clause of the US Constitution preempts states from imposing legal requirements (in this case, via a tort action in state court) that directly contradict federal law [52]. Driven by the political preferences of its leadership at the time, the FDA filed an amicus brief in

the case on behalf of the defendant manufacturer, arguing that imposition of product liability would "undermine the agency's authority to protect the public health" [53]. The brief claimed that an adverse court ruling would force companies to add warnings not approved by the FDA and could upset the delicate benefit/risk balancing that went into the construction of the drug labeling, which could result in overwarning and ultimately underuse of an effective drug.

The major deficiency in the logic of those favoring FDA preemption in this area is that these arguments inappropriately regard the FDA's official label as the final word on drug safety. In fact, preapproval clinical trials necessarily involve only a limited sample of patients and are often powered to detect changes in efficacy-related endpoints, rather than rates of adverse events (see Chapter 4). The FDA will not have a complete picture of the safety of drugs, even at the time the labeling is written. After approval, the FDA lacks the resources and capability to actively monitor evolving knowledge about a drug [54]. Until the FDA Amendments Act (FDAAA) of 2007 (Public Law 110-85), the FDA had no authority to compel manufacturers to update the warnings in drug labeling. After the withdrawal of rofecoxib, Sandra Kweder, Deputy Director of the FDA's Office of New Drugs, said in testimony at a US Senate hearing, "We don't have the authority to tell a company, 'This is how your label has to look. This is the language that needs to go into your label. Here is where it goes, end of story.' We have to negotiate with the company the specific language of how things should be worded, the placement, those kinds of things" [55]. The FDAAA gave the FDA limited authority to "require" labeling changes "if the Secretary becomes aware of new safety information that the Secretary believes should be included in the labeling of the drug," but made these decisions reviewable through an alternative dispute resolution procedure [56]. Although this new authority strengthened the FDA's hand somewhat, ensuring compliance can still involve a lengthy and resource-intensive legal process. While the pathway established by FDAAA has rarely been publicly invoked in the decade since passage of the law, its existence may strengthen the FDA's position in its negotiations with manufacturers over inclusion of warning language in a drug's labeling.

Manufacturers, by contrast, are in an optimal position to learn about emerging safety concerns after FDA approval, because they closely monitor the use of their products, organize postmarketing studies, and receive spontaneous reports from physicians and other sources about adverse events arising in the course of therapy (see Chapter 7). Manufacturers have a strong financial incentive to increase sales of their products, but manufacturers may also sometimes be faced with their own safety-related data that suggest limiting use of their product, or withdrawing it from the market altogether. In such situations, manufacturers have made poor decisions that adversely affect public health. For example, when drug safety issues have emerged after approval, some manufacturers have decided to downplay reports of side effects to physicians [57] and the FDA [58,59]. Failure-to-warn litigation, therefore, serves an important supplementary regulatory function – without undermining FDA requirements – by providing a disincentive (in the form of substantial monetary penalties) for manufacturers' decisions to hide or downplay reports of safety issues that emerge after a product reaches the market. Notably, former FDA commissioners have confirmed that "Although the FDA might later disapprove of a [strengthened warning] label …, the FDA's power to disapprove does not make the manufacturer's voluntarily strengthened label a violation of federal law" [60]. At any time, a manufacturer can strengthen the labeling by adding warnings to it without first notifying the FDA and receiving approval to do so. In fact, the Code of Federal Regulations states, "The labeling shall be revised to include a warning as

soon as there is reasonable evidence of an association of a serious hazard with a drug; a causal relationship need not have been proved" [61].

Despite these considerations, the FDA's amicus brief argument was repeated in subsequent failure-to-warn cases, and a few courts expressly adopted the position [62]. In 2006, the FDA attempted to solidify its position further in a surprise preamble to a set of regulations regarding the format of the label, in which it reiterated its new contention that any FDA-approved labeling, "whether it be in the old or new format, preempts … decisions of a court of law for purposes of product liability litigation" [63]. The FDA suggested that preemption should apply even if a manufacturer failed to warn adequately about a known risk, unless a patient could prove that the company intentionally committed fraud on the FDA, which is a very difficult legal standard to meet [64].

Ultimately, the US Supreme Court reviewed the legal foundation of the claimed FDA preemption of product liability related to prescription drugs. The pivotal case, *Wyeth v. Levine*, was based on a lawsuit from a patient who was treated with an intravenous antinausea medication for her migraine headache. The product extravasated and caused gangrene in her forearm, leading to amputation. The patient sued the drug manufacturer for inadequately warning on the label about the known risks of certain intravenous uses of its medication. A Vermont jury determined after fully considering the record that the label did not sufficiently describe the drug's known risks with intravenous drip administration. The manufacturer appealed the verdict, and the Vermont Supreme Court affirmed, finding that the jury's verdict did not conflict with the FDA's labeling requirements, which "create a floor, not a ceiling, for state regulation" [65].

The manufacturer appealed again to the Supreme Court, arguing that it was impossible to comply with the federally approved label and that the state court judgment would obstruct the purpose of federal drug laws. The manufacturer charged that the FDA, not the drug manufacturer, had the primary responsibility for the drug label. In a 6–3 decision, the Supreme Court upheld the Vermont decision and struck down the notion of federal preemption in this field [66]. Justice John Paul Stevens, writing for the majority, noted, "It has remained a central premise of drug regulation that the manufacturer bears responsibility for the content of its label at all times."

After *Wyeth v. Levine*, there remains no controversy about whether FDA approval of a drug label preempts failure-to-warn claims. However, the decision did leave open the possibility that preemption could be invoked if the FDA had "consider[ed] and reject[ed] a stronger warning." That is, if the FDA reviews all the data surrounding a particular safety issue and makes a specific statement that a strong warning is not necessary, such an action could be invoked by a defendant to support preemption of a failure-to-warn lawsuit.

Product Liability Law in Europe

The European Union (EU) is a political and economic coalition that currently consists of 28 countries in Europe. The main sources of EU law are regulations, directives, and decisions. Regulations are immediately enforceable in member states when they come into force and automatically override conflicting local provisions. By contrast, directives usually leave member states discretion as to how they are to be adopted [67].

Product liability has been called an "American invention" [68], and the general product liability directive for the EU (85/374/EEC) was originally enacted in 1985. A liability action arising under this directive includes the following contentions: (i) defective product; (ii) causation of damage' and (iii) no exclusion of liability.

A product is defective if it does not provide the safety that a person is entitled to expect, taking all circumstances in account, including the presentation of the product, the use reasonably expected of the product, and the time when the product was put into circulation. However, a product may not be considered defective simply because a better product was subsequently put into circulation [69].

Product liability in non-EU European countries is defined by country-specific laws, although the approach is similar to that in the EU. In Switzerland (a non-EU country), for example, there is no specific legislation covering drug liability. Nonetheless, the generally applicable Product Liability Act requires a showing of (i) damage (design defect or failure to warn); (ii) causation of damage; and (iii) no exclusion of liability.

Like in the US, most product liability lawsuits in EU and non-EU countries in Europe are based on failure-to-warn claims about the adverse event at issue, rather than design defects. One of the exclusions of liability, as in US, is the learned intermediary defense. For example, in Switzerland, a 16-year-old patient was prescribed the oral contraceptive medication drospirenone/ethinylestradiol (Yasmin) and experienced a thromboembolism with a subsequent stroke. The patient sued the manufacturer based on a failure-to-warn claim. The labelling states that there is a higher risk for thrombosis and in the information letter to the physician, the risk for a thromboembolism is reported to double with the intake of the contraceptive medication. The Supreme Court ruled that this information was sufficient [70]. As previously discussed, since DTCA is not available in Europe, it is not possible for that practice to undermine the learned intermediary defense.

Within the EU, because the product liability rule is a directive, member states retain some flexibility in implementing aspects of it, such as whether they permit compensation for noneconomic damages (e.g., pain and suffering) or which manufacturer defenses they seek to incorporate [71]. As a result of this flexibility, there is substantial diversity across EU countries in how product liability cases are adjudicated [72]. Still, countries can be limited by the directive. For example, in a series of cases, the European Court of Justice prevented France, Spain, and Denmark from enacting provisions considered to be too friendly to the damaged party [73].

Product liability related to prescription drugs in Germany bears special attention, since this is the only EU country that has implemented (even prior to the enactment of the EU liability directive) particular rules in this field via its Medicines Act. This is the consequence of the international thalidomide birth defect public health crisis of the late 1950s and early 1960s, which affected approximately 7000 children in Germany alone [74]. Liability exists if, when used in accordance with the intended purpose, the drug has harmful effects which exceed the limits considered tolerable in the light of current medical knowledge (i.e., a design defect), or the damage has occurred as a result of labelling that does not comply with current medical knowledge (i.e., a failure to warn) [75]. However, determining liability is based on a strict liability model that requires only demonstration of (i) damages; (ii) causation of damages; and (iii) no exclusions from liability (e.g., the learned intermediary defense) [76]. Another characteristic of German drug liability is the limitation of the amount of compensation for damages. In a case of death of or injury to a person, the pharmaceutical company is liable only for a capital amount of up to €600 000 or an annuity of up to €36 000 per year. In a case of death of or injury to several persons by the same drug, the pharmaceutical company shall be liable for a capital amount of up to €120 million or an annuity of up to €7.2 million per year [77].

German law also provides an interesting case example about the labeling requirements covering prescription drugs in Europe. The European Medical Agency (EMA) regularly

issues guidelines that specify the content and presentation of the labeling [78], including for example categorizing the frequency of possible side effects as very often (>10%), often (1–10%), occasional (0.1–1%), rare (0.01–0.1%), and very rare (<0.01%). In addition, Section 10 of the German Medicines Act lists all categories of information that need to be placed on a drug's labeling, such as indications, dosage, duration of intake, reference to overdose, expiration date, and adverse events. In describing adverse events, manufacturers must include a description of all adverse reactions that can occur when the drug is used as intended, the countermeasures to be taken if possible in the event of adverse reactions, and an additional standard text that explicitly instructs patients to inform their physicians, pharmacists, health professionals, or the competent higher federal authority directly of every suspected adverse reaction [79]. The German Supreme Court has ruled that more detailed information must be provided as the severity and probability of a potential adverse event increase [80].

Overall, product liability law in Europe is in many ways similar to that in the US, especially with regard to the principles of strict liability and the learned intermediary defense. However, failure-to-warn claims are less likely to succeed in Europe than in the US, and damages practices and rules generally lead to lower compensation for patients. As a result, fewer drug liability claims are brought in Europe, and the outcome of a case can vary widely whether a claim is being brought to court in the US or in a European country.

Pharmacoepidemiology and Contract Law

Many studies in the field of pharmacoepidemiology emerge from collaborations among individuals at different institutions. Different researchers may bring specific types of expertise to a project or different resources [81,82]. For example, researchers may have all the computing power they need, but require access to a certain external database to address a question. Collaborations may occur among academic centers, between nonprofit and for-profit companies, or with the government. Cooperative work can allow more complex research to be performed and help advance the field of pharmacoepidemiology in several ways.

One type of collaborative work of particular public health importance is contract research. Contract research is undertaken by an individual, academic, or nonprofit investigator supported by a sponsor (usually an industry or governmental agency). Most contractual research relationships are defined by the generation of a "deliverable," which can be a database, a research report, or some other product. The contract is the centerpiece of the relationship and classically represents the full outline of the agreement between the parties. The mutually agreed-upon terms are used as evidence of the parties' intentions if the agreement later runs into trouble and ends up in court. Relationships with industry are common; one survey of clinical epidemiologists and health services researchers in the US found that about 40% reported currently being involved in such relationships, while 50% reported forming collaborations with industry leading to publications [83]. In countless cases, contract research in pharmacoepidemiology has led to important public health findings and changes in healthcare delivery.

However, contract research may pose various potential pitfalls as well. Concern about contract research generally centers around (i) trial design; (ii) access to data and data analysis; and (iii) publication of results. It has long been known that there is a statistically significant relationship between a favorable study result and the source of research funding [84,85]. These results can be explained by choices made in trial design, when subjective decisions about

comparators [86] or the inclusion or exclusion of certain variables or potential confounders in epidemiologic and economic studies can affect the ultimate results of the trial [87]. Investigators should be wary of performing contract research in which the sponsor has the right to unduly influence the design of the trial. Many sponsors prefer to retain control of the data and insert their own statistical analyses. They argue that such efforts guard against "investigators [who] want to take the data beyond where the data should go," while investigators argue that this arrangement provides the company with an opportunity to "provide the spin on the data that favors them" [88]. In one case of an experimental AIDS vaccine, after a negative trial, the sponsor demanded that its contradictory analyses be inserted into the manuscript and ultimately sued the investigators for $7 million after the article was published [89].

Access to clinical trial data is critically important for academic researchers. In the case of rosiglitazone, a clinical trial organized by the manufacturer sought to compare the product against other treatment options for diabetes, and an independent academic steering committee was organized to oversee the data analysis [90]. Company documents suggest that the clinical trial database was exclusively controlled by the company, which provided limited access to the investigators [91]. When members of the steering committee questioned the presentation of the results, their concerns were largely overlooked [77]. In reviewing this case, one commentator concluded that the absence of independent access to all of the data in the trial may allow physician-scientists to be manipulated by the sponsor, resulting in a manuscript that does not provide the most accurate assessment of the risks and benefits of the therapy [77]. Contracts should be carefully scrutinized for the way in which they delineate who controls access to the data.

Finally, there have been conflicts over so-called gag clauses that prevent contract investigators from publishing their results [92]. For example, when a University of Toronto physician identified safety issues related to an experimental drug used to treat iron overload in transfusion-dependent patients with thalassemia [93], she was not granted permission to publish her results. When she ultimately exposed her findings, she was the subject of a breach of contract lawsuit from the sponsor, on the basis that her research contract provided that the published work-product was "secret and confidential" and could not be disclosed except with the manufacturer's "prior written consent" [94]. In the case of the cholesterol-lowering drug ezetimibe (Zetia®), the outside investigator leading a large-scale clinical trial found that the drug lacked important efficacy in cardiovascular outcomes. He reportedly pressured the manufacturer to no avail to speed the release of the data, and due to contractual obligations was unable to come forward with the data on his own without such approval [95].

Such problems are not limited to private industry contracts. In the US, a report from the Association of American Universities and the Council on Government Relations found that federal agencies commonly include controls on the dissemination of research results in their sponsored contracts and grants [96]. Contracting issues related to liability, trial design, access to data and data analysis, and publication of results are also not limited to a particular country [97]. In Europe, for example, countries aware of the challenges in setting up contracts between investigators and industry in particular offer government assistance and templates to help balance the diverging interests. In Switzerland, the ethics committee provides templates for clinical trial agreements on its website [98], and the UK and the European Commission also offer such templates and guidance notes [99].

For researchers based in academic medical centers, institutional research administration offices usually handle the details of contract

negotiation with research sponsors. However, a survey of academic medical centers in 2001 found that academic institutions routinely engage in industry-sponsored research without sufficient protection for investigators [100]. For example, a median of 1% of research administration offices (interquartile range 0–21%) in US universities reported requiring that authors have access to all the data for multicenter trials. A 2005 survey found little change. Nearly half of academic institutions reported that they allowed contract provisions permitting the research sponsor to insert its own statistical analyses and draft the manuscript, while prohibiting investigators from sharing data with third parties after a trial had ended. The survey also found that 17% of academic research centers reported disputes between researchers and sponsors about control of or access to data [101].

A few expert bodies have offered recommendations on legal guidelines for the conduct of contract research [102]. The best known and most authoritative have emerged from the International Committee for Medical Journal Editors (ICMJE). Their guidelines for original research articles submitted to biomedical journals require that the investigators be independent of the sponsors' role in the research, fully accountable for the design and conduct of the trial, have independent access to all trial data, and control all editorial and publication decisions [103]. Each of these criteria must be worked out at the beginning of the contractual relationship between the sponsor and investigators.

Whether or not they receive support from research administration offices, pharmacoepidemiologists must be aware of the ICMJE guidelines and thoroughly evaluate contracts guiding research for inappropriate language regarding control of design of the trial, access to data, and reporting of results (see Table 9.2). They should also be aware that some peer-reviewed journals have even more strict standards than the ICMJE; for example, *Pharmacoepidemiology and Drug Safety* currently requires disclosure of any control the sponsor had on the study and manuscript. Problematic language includes overly broad confidentiality clauses, clauses that define and assign ownership of intellectual property,

Table 9.2 Potentially objectionable language in research contracts for pharmacoepidemiologists.

Category	Contractual terms	Critique
Control over investigator work-product	"____ shall provide confidential information to CONSULTANT for the purpose of conducting the CONSULTANT'S professional services. All information whether written or verbal provided by, or developed for _____, and all data collected during the performance of this Agreement is deemed to be the Confidential Information of _____."	Broad definition of "confidential information" seems to cover all information. Researcher's work-product becomes sponsor's confidential information.
Gag clauses	"No information regarding this Agreement or the interest of ____ or Client in the subject matter hereof shall be disclosed to any third party without the prior written consent of _____"	Prevents disclosure of existence of the contract as a financial source in publication.
Opportunity to influence outcome	Client "shall not present or publish, nor submit for publication, any work resulting from the Services without _____ prior written approval."	Contract allows sponsor to quash publication unless it approves analyses.

All examples are anonymized but otherwise unchanged excerpts from actual contracts written to cover sponsored pharmacoepidemiologic research.

and clauses that require approval from a sponsor prior to publication. It may be reasonable to allow sponsors a limited amount of time to review proposed publications for inadvertent release of proprietary company information or to contribute suggestions based on their expertise. However, researchers have an ethical obligation to ensure that contracts do not unreasonably delay the publication of potentially important results. Poorly written contracts can lead to inappropriate secrecy of results, which can have public health concerns, as well as resulting in litigation against researchers. Balancing the contractual tightrope might not be easy, but it is important. As Dr. Curt Furberg has said, "Companies can play hardball, and many investigators can't play hardball back. You send the paper to the company for comments, and that's the danger. Can you handle the changes the company wants? Will you give in a little, a little more, then capitulate? It's tricky for those who need money for more studies" [104].

Pharmacoepidemiology and Intellectual Property Law

Patent law is a field of growing importance to the practice of pharmacoepidemiology. A patent is a formal grant of market exclusivity authorized by the federal government. The concept of a patent may have originated in ancient Greece, but became a formal legal instrument in England and Europe in the fourteenth and fifteenth centuries. In the US, the original Patent Act was passed under authority from the Constitution, which permits Congress to develop laws that "promote progress of Science and the Useful Arts" [105]. Patents give inventors the right to exclude others from making, using, offering to sell, or selling the invention claimed in the patent for 20 years from the patent application date [106]. The goal of a patent is to encourage inventors to invest in the development of their ideas, because it gives them a competition-free period in which to market a successful invention. Patents can be issued for any process, machine, manufacture, or composition of matter. To be worthy of a patent, an innovation in one of these categories must be useful, novel, and nonobvious to a person of ordinary skill in the field. These criteria aim to ensure that patents cannot be awarded for inventions that already exist, or small, noninnovative improvements on those inventions. In recent years, numerous patents have been obtained on methods and techniques used in pharmacoepidemiology, including investigating characteristics of drug use and adverse events.

In filing for a patent, an inventor must fully disclose the content of the claimed invention in a patent document. This disclosure must provide clear detail about the invention and must enable any person skilled in the art to use it, including the "best mode" (if they have contemplated one) available for making the inventions work. The process for obtaining a patent involves submitting the patent document to examiners at institutions such as the European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO) who have expertise in the general subject matter of the patent. An examiner checks the application for technical accuracy and evaluates the innovativeness of the claimed invention by comparing it to previous publications and issued patents (in legal terminology, publicly available documents such as these are termed the "prior art"), to see if all the basic criteria are met. This process generally involves substantial back-and-forth between the examiner and the applicant, and may take several years to complete. Inventors may submit patent applications themselves, or enlist the help of specially trained patent agents or patent attorneys.

Inventors may have numerous justifications for pursuing patents. First, patents provide an incentive for investment in research by offering an opportunity to recoup start-up costs after

dissemination of a product. Other inventors may seek a way to publish their innovative processes while still retaining control over what they consider to be their intellectual property. A patent is classically thought of as a "quid pro quo" between inventors and society [107]. The government provides its police power to protect an inventor's intellectual property for a set length of time and, in exchange, the inventor makes the invention available to the public and fully describes it, so that others can use it and potentially improve on it in subsequent innovation. However, patents can also be controversial. Patents over scientific research tools have been implicated in barriers to effective cooperation [108], enhanced secrecy among researchers [109], and restrictions on availability of the products of research to patients [110].

Patents have become increasingly visible in the practice of pharmacoepidemiology. Most fall into the "process" category, such as methods of analyzing claims data and comparing outcomes to identify adverse events. The US Supreme Court has held that patentable processes may not include fundamental principles such as "laws of nature, natural phenomena, or abstract ideas" [111], or purely mental processes [112]. However, applications of laws of nature to a particular process may still be patentable. For example, a well-known case involved a patent over a method of curing synthetic rubber that used the Arrhenius equation to calculate the optimal cure time. The process was found to be patentable because the formula was a part of a larger inventive process for curing rubber [94].

Patents related to the practice of pharmacoepidemiology have been obtained by applicants ranging from individuals (e.g., a patent covering a method for assessing the association of genomic data with drug safety adverse event data [113]) to large healthcare data collectors such as Microsoft (e.g., a patent covering a method for large-scale data collection and data mining to infer health-related observations [114]). For example, one patent was awarded to inventors and assigned to a start-up company for a "method, system, and software for analyzing pharmacovigilance data." The patent covers a process of:

> [D]etermining a sample size-independent measure of association between two conditions of interest in the dataset of pharmacovigilance data; using a hypergeometric distribution to determine a measure of statistical unexpectedness between the conditions of interest in said dataset … and displaying the measure of association with the measure of the statistical unexpectedness to identify a significant association between conditions of interest. [115]

The concept of "hypergeometric distribution" may not be patentable as an abstract idea, but in this case the USPTO clearly considered the process patentable overall despite its integral use of that principle.

There are important ethical and legal concerns related to patenting processes that provide exclusive control over various aspects of the conduct of pharmacoepidemiology and pharmacovigilance research. First, patents that are sufficiently broad could prevent others from conducting necessary research into drug outcomes and effects, unless potentially expensive third-party licenses were negotiated beforehand. In one case, an HIV researcher at Stanford faced a patent-infringement lawsuit over a publicly available database he created to help guide antiretroviral therapy based on the resistance characteristics of the disease, because searching this database may involve a similar process to one previously patented (but never implemented) by a for-profit company [116]. In another case, a patent-seeker in the field argued that researchers should patent the adverse reactions discovered in pharmacoepidemiologic studies to enhance funding from for-profit pharmaceutical companies that might be interested in novel and nonobvious processes that link drugs and adverse events [117]. However, a

proliferation of patents over processes linking drug delivery to reported adverse events could increase costs through "another layer of bureaucrats and patent attorneys" and hurt the public health, as "real information could get easily lost in a blizzard of patented factoids" [118].

The US Supreme Court has stepped into the controversy over process patents. In 2008, the Court of Appeals for the Federal Circuit, the highest US patent appeals court below the Supreme Court, revisited its interpretation of what may be considered a patentable process. The case involved a patent over a business method for reducing risk in situations of fluctuating prices. The Federal Circuit Court held that for a process to be patentable, it must be tied to a particular machine or apparatus, or transform an object into a different state or thing [119]. Notably, as pertaining to pharmacoepidemiologic patents, the Federal Circuit Court held that "in most cases, gathering data would not constitute a transformation" because "every algorithm inherently requires the gathering of data inputs" [120]. The Supreme Court in *Bilski v. Kappos* reviewed this standard and agreed that the machine-or-transformation test was one valid way of determining whether a business method was patentable, although it was not the exclusive test [121].

Despite the Supreme Court's reluctance to draw a bright line separating patentable from nonpatentable processes, the Court's support for the machine-or-transformation test may undercut certain patents related to pharmacoepidemiology and pharmacovigilance [122]. For example, the Federal Circuit Court used the test to invalidate a patent related to a method of adverse effect detection [123]. In that case, an inventor had secured a patent on a method of using adverse event data regarding vaccine administration to inform subsequent healthcare delivery. The patent at issue claimed:

> A method of determining whether an immunization schedule affects the incidence or severity of a chronic immune-mediated disorder in a treatment group of mammals, relative to a control group of mammals, which comprises immunizing mammals in the treatment group of mammals with one or more doses of one or more immunogens, according to said immunization schedule, and comparing the incidence, prevalence, frequency, or severity of said chronic immune-mediated disorder or the level of a marker of such a disorder, in the treatment group, with that in the control group. [124]

Pharmacoepidemiologists are likely to continue to come across patented methods in their daily work and be faced themselves with the question of whether to pursue patents on their research tools. This is particularly true in the US, where the *Bilski* decision left the door open for patents to be issued on processes involved in medical practice or pharmacoepidemiologic research.

Intellectual Property Law in Europe

In Europe, the European Patent Convention (EPC) provides the legal framework under which patents are granted. The establishment of patentability is framed in terms of fulfilling three prerequisites: novelty, usefulness, and an inventive step (equivalent to the "nonobviousness" requirement under US law) [125]. Like in the US, the maximum term of a European patent is 20 years from its filing date [126].

The US and European standards with regard to the patentability for methods and techniques are also close. The EPC provides a nonexhaustive list of nonpatentable inventions: discoveries, scientific theories, and mathematical methods; aesthetic creations; schemes, rules, and methods for performing mental acts, playing games, or doing business, and programs for computers; and presentations of information

[127]. Patents can be obtained on software according to the Technical Boards of Appeal of the EPO if the software produces a further technical effect when it runs on a computer

> which goes beyond the "normal" physical interactions between program (software) and computer (hardware). … Although it may be said that all computer programming involves technical considerations since it is concerned with defining a method which can be carried out by a machine, that in itself is not enough to demonstrate that the program which results from the programming has technical character; the programmer must have had technical considerations beyond "merely" finding a computer algorithm to carry out some procedure. [128]

According to the Guidelines for Examination in EPO, such a further technical effect can be found, for example, in the control of an industrial process or in the internal functioning of the computer itself or its interfaces under the influence of the program, or can affect the efficiency or security of a process. Software that implements a mathematical method that itself makes a technical contribution can also qualify as a further technical effect [129].

There are three possible routes for obtaining patent protection in Europe. One can apply for a patent directly to the national patent office of a particular country (national patent); one can apply for a patent to the EPO and designate specific EU member states where patent protection is wanted ("classical" European patent); or – as part of a new pathway intended to start in 2019 – one can apply for a patent to the EPO with the designation of a unitary patent that will be applicable for all of the EU member states where the government has ratified the Agreement on a Unified Patent Court [130].

The European patent system enables a central examination by the EPO, which is more efficient than the national patent process. However, granted European patents have to be subsequently validated individually in each country in which they are intended to take effect, and validation requirements can differ. The goal of the new unitary patent system is to reduce complexity and lower costs. Unitary patents will confer uniform protection, since the substantive patent law has been harmonized in the Agreement on a Unified Patent Court [131], which 25 EU member states have ratified (up to 2017) [132]. The member states also set up a Unified Patent Court to deal with the infringement and validity of unitary patents and European patents, intended to enhance legal certainty through harmonized case law in the area of patent infringement and validity and enable more efficient judicial procedures [133].

The choice among seeking a national patent, European patent, or unitary patent needs to be made depending on the preferences of the individual applicant. For example, applicants should weigh the need for broad geographic coverage versus protection in one (or a few) member states. Furthermore, consideration should also be given to whether the patent should be subject to the exclusive jurisdiction of the Unitary Patent Court, or if it is preferred to use national courts with a more limited geographic jurisdiction. While a classical European patent contains the costs for validation and renewal fees in each member state in which protection is required, the unitary patent does not include validation costs, except the cost for one translation during the transitional period as well as a single renewal fee [134].

Conclusion

Legal issues intersect with the practice of pharmacoepidemiology in many ways. Pharmacoepidemiologists may be involved in product liability cases brought by individuals against drug manufacturers, either as expert witnesses or on the basis of academic work they

undertake. These cases traditionally involve a claim of a failure to warn, which requires proof that the manufacturer knew of the safety issue, that any provided warnings were insufficient, and that the injury received was directly caused by use of the drug. Manufacturers can invoke a "learned intermediary" defense to deflect responsibility onto the treating physician, but in the US after *Wyeth v. Levine* can no longer argue that FDA approval of the drug labeling precludes providing additional warnings about adverse effects for cases in which the warnings are warranted by the data. While similar product liability rules apply in Europe, fewer cases are brought to court and damage compensation is lower.

Pharmacoepidemiologists may also be involved in contract research, but should carefully consider contractual requirements related to ownership of the work product and withholding publication.

Finally, both in the US and in Europe, pharmacoepidemiologists may decide to try to patent their research methods, but should weigh up the risks and benefits of this form of intellectual property.

References

1 Kesselheim AS, Avorn J. The role of litigation in defining drug risks. *JAMA* 2007; **297**: 308–11.
2 *Thomas v. Winchester*, 6 N.Y. 397 (N.Y. 1852).
3 Restatement (Second) of Torts § 402A (1965).
4 Restatement (Second) of Torts § 402A (1965) comment j.
5 Restatement (Second) of Torts § 402A (1965) comment k.
6 Prosser WL. *Handbook of the Law of Torts*, 5th edn. St. Paul, MN: West Publishing 1984, § 99.
7 *Brown v. Superior Court (Abbott Labs)*, 751 P.2d 470 (Cal. 1988).
8 *Feldman v. Lederle*, 479 A.2d 374 (N.J. 1984).
9 Restatement (Third) of Torts: Products Liability § 6c (1997).
10 Fischer DA, Powers W. *Products Liability: Cases and Materials*. St. Paul, MN: West Publishing, 1988.
11 *Chambers v. G.D. Searle & Co*., 441 F. Supp. 377 (D. Md. 1975).
12 *Spitzer v. GlaxoSmithKline*, Civ Action #04-CV-5304, S.D.N.Y. 26 Aug 2004.
13 Kondro W, Sibbald B. Drug company experts advised staff to withhold data about SSRI use in children. *CMAJ* 2004; **170**(5): 783.
14 Olfson M, Gameroff MJ, Marcus SC, Waslick BD. Outpatient treatment of child and adolescent depression in the United States. *J Am Acad Child Adolesc Psychiatry* 2001; **40**: 762–72.
15 Food and Drug Administration. Suicidality in children and adolescents being treated with antidepressant medications. 2004. https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/suicidality-children-and-adolescents-being-treated-antidepressant-medications (accessed May 2019).
16 In Re: Bayer AG Securities Litigation. 2004 US Dist LEXIS 19593 (S.D.N.Y., 2004).
17 Berenson A. Trial lawyers are now focusing on lawsuits against drug makers. *NY Times*, 18 May 2003.
18 *Delaney v. Warner-Lambert Co*., 2003 Cal. App. Unpub. LEXIS 2883 (Cal Ct App., 26 Mar 2003).
19 Watkins PB, Whitcomb RW. Hepatic dysfunction associated with troglitazone. *New Engl J Med* 1998; **338**: 916–17.
20 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New Engl J Med* 2007: **356**; 2457–71.

21 Staff Report on GlaxoSmithKline and the Diabetes Drug Avandia. Prepared by the staff of the Committee on Finance, United States Senate, Max Baucus, Chairman, Charles Grassley, Ranking Member. https://www.finance.senate.gov/release/grassley-baucus-release-committee-report-on-avandia (accessed May 2019).

22 Harris G, Koli E. Lucrative drug, danger signals and the FDA. *NY Times*, 10 Jun 2005.

23 Bombardier C, Laine L, Reicin A, *et al*. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *New Engl J Med* 2000; **343**(21): 1520–8.

24 *McDarby v. Merck*, 949 A.2d 223 (N.J. Super. Ct, 2008).

25 *Barson v. E.R. Squibb & Sons*, 682 P.2d 832 (Utah 1984).

26 Brennan TA. Causal chains and statistical links: the role of scientific uncertainty in hazardous-substance litigation. *Cornell Law Rev* 1988; **73**: 469–533.

27 *Siharath v, Sandoz*, 131 F. Supp. 2d 1347 (N.D. Ga., 2001).

28 Federal Judicial Center. *Reference Manual on Scientific Evidence*, 2nd edn. Washington, DC: FDC, 2000. https://www.fjc.gov/sites/default/files/2012/sciman00.pdf (accessed May 2019).

29 *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311 (9th Cir., 1995).

30 *Allison v. McGhan Medical Corp*, 184 F.3d 1300 (11th Cir., 1999).

31 Ross JS, Madigan D, Hill KP, Egilman DS, Wang Y, Krumholz HM. Pooled analysis of rofecoxib placebo-controlled clinical trial data: lessons for postmarket pharmaceutical safety surveillance. *Arch Intern Med* 2009; **169**(21): 1976–85.

32 *Merck v. Garza*, No. 04-07-00234-CV (Texas Ct App, 14 May 2008).

33 Kesselheim AS, Studdert DM. The Supreme Court, preemption, and malpractice liability. *New Engl J Med* 2009; **360**: 559–61.

34 Restatement (3rd) of Torts: Products Liability §6 (1998).

35 *Carlin v. Superior Court*, 13 Cal. 4th 1104 (1996).

36 *Martin v. Hacker*, 586 N.Y.S.2d 407 (N.Y. App. Div., 1992).

37 *Perez v. Wyeth Labs*, 734 A.2d 1245 (N.J. 1999).

38 *State ex rel. Johnson & Johnson v. Karl*, 647 S.E.2d 899 (W.Va., 2007).

39 Kesselheim AS, Studdert DM. Role of medical professional organizations in regulating physician expert witness testimony. *JAMA* 2007; **298**: 2907–9.

40 *Daubert v. Merrell Dow Pharmaceuticals*, 509 US 579 (1993).

41 *General Elec. Co. v. Joiner*, 522 US 136 (1997).

42 *Rosen v. Ciba-Geigy Corp.*, 78 F.3d 316 (7th Cir. 1996).

43 Federal Rules of Evidence § 702.

44 *Frye v. United States*, 293 F. 1013 (D.C. Cir., 1923).

45 Robertson CT. Blind expertise: a litigant-driven solution to expert bias. *NYU Law Rev* 2010; **85**: 174–257.

46 Gatowski SI, Dobbin SA, Richardson JT, Ginsburg GP, Merlino ML, Dahir V. Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-*Daubert* world. *Law Human Behav* 2001; **25**(5): 433–58.

47 21 USC § 355 (2009).

48 21 USC § 355(b)(1)(F) (2009).

49 Kesselheim AS, Avorn J. Pharmaceutical promotion to physicians and First Amendment rights. *N Engl J Med* 2008; **358**: 1727–32.

50 Kessler DA, Vladeck DC. A critical examination of the FDA's efforts to preempt failure-to-warn claims. *Geo Law J* 2008; **96**: 461–95.

51 *Motus v. Pfizer*, 358 F.3d 659 (9th Cir., 2004).

52 Kesselheim AS. Permitting product liability litigation for FDA-approved drugs and devices promotes patient safety. *Clin Pharmacol Ther* 2010; **87**(6): 645–7.

53 Brief for *Amicus Curiae the United States of America, Motus v. Pfizer*, Nos. 02-55372, 02-55498 (9th Cir.), filed 3 Sep 2002.

54 Government Accountability Office. High-risk series: an update. GAO Report GAO-09-271. 2009. https://www.gao.gov/new.items/d09271.pdf (accessed May 2019).

55 Kweder S, Deputy Director Office of New Drugs, Food and Drug Administration. FDA's drug approval process: up to the challenge? Hearing before the Senate Health, Education, Labor & Pensions Committee, 109th Cong. 1 Mar 2005. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=109_senate_hearings&docid=99-761 (accessed May 2019).

56 21 USC §355(o)(4)(a) (2009).

57 Berenson A. For Merck, Vioxx paper trail won't go away. *NY Times*, 21 Aug 2005.

58 Psaty BM, Kronmal RA. Reporting mortality findings in trials of rofecoxib for Alzheimer disease or cognitive impairment. *JAMA* 2008; **299**: 1813–17.

59 Psaty BM, Furberg CD, Ray WA, Weiss NS. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: use of cerivastatin and risk of rhabdomyolysis. *JAMA* 2004; **292**: 2622–31.

60 Brief of *amici curiae* Donald Kennedy, M.D., and David A. Kessler, M.D., in support of respondent. Supreme Court of the United States 06-1249. 14 Aug 2008.

61 21 CFR 210.57(e).

62 *Colacicco v. Apotex*, 521 F. 3d 253 (3d Cir, 2008).

63 Food and Drug Administration. Requirements on content and format of labeling for human prescription drug and biologic products. *Fed Regist* 2006; **71**: 3922–97.

64 Avorn J, Shrank W. Highlights and a hidden hazard: the FDA's new labeling regulations. *N Engl J Med* 2006; **354**: 2409–11.

65 *Wyeth v. Levine*, 944 A.2d 179 (Vt. 2006).

66 *Wyeth v. Levine,* 129 S. Ct. 1187 (2009).

67 Jaag T, Hänni J. *Europarecht: die europäischen Institutionen aus schweizerischer Sicht*. Zurich: Schultess, 2015.

68 Mann LC, Rodrigues PR. The European Directive on Products Liability: the promise of progress? *Ga J Int Comp Law* 1988; **18**: 391–426.

69 Art. 6 785/374/EEC.

70 Supreme Court decision from January 5, 2015 (4A_365/2014; 4A_371/2014).

71 E.g., Art. 7 85/374/EEC.

72 Magnus U. Das deutsche Deliktsrecht im europäischen Kontext, *HAVE* 2017; **1**: 25–37.

73 ECJ Slg. 1993, I-1 (Rs. C-293/91, *Commission./.France*); ECJ Slg. 1997, I-2649 (Rs. C-300/95, *Commission./.United Kingdom*); ECJ Slg. 2002, I-3827 (Rs. C-52/00, *Commission./.France*); ECJ Slg. 2002, I-3879 (Rs. C-154/00, *Commission./.Greece*); ECJ 14.3.2006 (Rs. C-177/04, *Commission./.France*).

74 Alto Charo, R. Speed versus safety in drug development. In Lynch HF, Cohen IG, ed., *FDA in the 21st Century: The Challenges of Regulating Drugs and New Technologies*. New York: Columbia University Press, 2015, p. 251.

75 § 84(1) Medicines Act (Arzneimittelgesetz).

76 § 84 Medicines Act (Arzneimittelgesetz).

77 § 88 Medicines Act (Arzneimittelgesetz).

78 European Medicines Agency. Product-information requirements. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000199.jsp (accessed May 2019).

79 § 10(5) Medicines Act (Arzneimittelgesetz).

80 BGHZ 106, 273 (Asthma-Spray); BGH VersR 1992, 96.

81 Kong SX, Wertheimer AI. Outcomes research: collaboration among academic researchers, managed care organizations, and pharmaceutical manufacturers. *Am J Manag Care* 1998; **4**(1): 28–34.

82 Eichler HG, Kong SX, Grégoire JP. Outcomes research collaborations between third-party payers, academia, and pharmaceutical manufacturers: what can we learn from clinical research? *Eur J Health Econ* 2006; **7**(2): 129–35.

83 Zinner DE, Campbell EG. Life-science research within US academic medical centers. *JAMA* 2009; **302**: 969–76.

84 Davidson RA. Source of funding and outcome of clinical trials. *J Gen Int Med* 1986; **1**: 155–8.

85 Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000–2005. *JAMA* 2006; **295**: 2270–4.

86 Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug–drug comparisons: why some statins appear more efficacious than others. *PLoS Med* 2007; **4**: e184.

87 Friedberg MA, Saffran B, Stinson TJ, Nelson W, Bennett CL. Evaluation of conflicts of interest in economic analyses of new drugs used in oncology. *JAMA* 1999; **282**: 1453–7.

88 Bodenheimer T. Uneasy alliance: clinical investigators and the pharmaceutical industry. *N Engl J Med* 2000; **342**: 1539–44.

89 Gottlieb S. Firm tried to block report on failure of AIDS vaccine. *BMJ* 2000; **321**: 1173.

90 Home PD, Pocock SJ, Beck-Nielsen H, *et al*. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet* 2009; **373**: 2125–35.

91 Nissen SE. Setting the RECORD straight. *JAMA* 2010; **303**(12): 1194–5.

92 Steinbrook R. Gag clauses in clinical-trial agreements. *N Engl J Med* 2005; **352**: 2160–2.

93 Olivieri NF, Brittenham GM, McLaren CE, *et al*. Long-term safety and effectiveness of iron-chelation therapy with deferiprone for thalassemia major. *N Engl J Med* 1998; **339**: 417–23.

94 Shuchman M. Legal issues surrounding privately funded research cause furore in Toronto. *CMAJ* 1998; **159**: 983–6.

95 Berenson A. Accusations of delays in releasing drug results. *NY Times*, 1 Apr 2008.

96 Association of American Universities/ Council on Government Relations. Restrictions on research awards: troublesome clauses 2007/2008. 2008. https://www.aau.edu/sites/default/files/AAU%20Files/Key%20Issues/Science%20%26%20Security/Report_AAU-COGR_Restrictions-on-Research-Awards-Troublesome-Clauses_2004.pdf (accessed May 2019).

97 Gisi D. Forschungsverträge und Patente: was gibt es zu beachten? CTU Lectures, Inselspital Bern. 2016. http://www.ctu.unibe.ch/unibe/portal/fak_medizin/micro_ctu/content/e182527/e312629/e422381/files422386/CTU2016Gisi_Forschungsvertrgefinal_eng.pdf (accessed May 2019).

98 Swiss Ethics Committee on research involving humans. Templates. http://www.swissethics.ch/templates_e.html (accessed May 2019).

99 European Commission. Research and innovation. https://ec.europa.eu/research/health/pdf/infoday_2015/template.pdf and http://www.ukcrc.org/regulation-governance/model-agreements (accessed May 2019).

100 Schulman KA, Seils DM, Timbie JW, et al. A national survey of provisions in clinical-trial agreements between medical schools and industry sponsors. *N Engl J Med* 2002; **347**: 1335–41.

101 Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *N Engl J Med* 2005; **352**: 2202–10.

102 Halperin SA, Scheifele D, Duval B, *et al*. Canadian Association for Immunization Research and Evaluation (CAIRE) guidelines for industry-sponsored clinical trial and epidemiologic contract research. *Hum Vaccin* 2005; **1**(4): 140–2.

103 International Committee of Medical Journal Editors. Uniform requirements for

manuscripts submitted to biomedical journals; writing and editing. 2018. http://www.icmje.org/icmje-recommendations.pdf (accessed May 2019).
104 Bodenheimer T. Uneasy alliance: clinical investigators and the pharmaceutical industry. *N Engl J Med* 2000; **342**: 1539–44.
105 U.S. Const. art. I, § 8, cl. 8.
106 35 U.S.C. § 101.
107 Porcelli FP, Dragseth JA. *Patents: A Historical Perspective*. Minneapolis, MN: Fish & Richardson, 2002.
108 Heller MA, Eisenberg RS. Can patents deter innovation: the anticommons in biomedical research. *Science* 1998; **280**: 698–701.
109 Blumenthal D, Campbell EG, Gokhale M, *et al*. Data withholding in genetics and the other life sciences: prevalences and predictors. *Acad Med* 2006; **81**: 137–45.
110 Merz JF, Kriss AG, Leonard DGB, Cho MK. Diagnostic testing fails the test. *Nature* 2002; **415**: 577–9.
111 *Diamond v. Diehr*, 450 U.S. 175 (1981).
112 *Gottshalk v. Benson*, 409 U.S. 63 (1972).
113 US Patent 7,461,006.
114 US Patent 7,406,453.
115 US Patent 7,650,262.
116 Mullin J. The fight of his life. *IP Law Bus* 2009; **Apr–May**: 44–8.
117 Classen JB. Enhanced funding of pharmacoepidemiology through patenting the disclosure of adverse event information. *Pharmacoepidemiol Drug Saf* 2006; **15**: 390–3.
118 Walker AM. More lawyers, more bureaucrats, less information on drug safety. *Pharmacoepidemiol Drug Saf* 2006; **15**: 394–5.
119 In re Bilski, 545 F.3d 943 (Fed. Cir. 2008) (en banc).
120 In re Bilski, 545 F.3d 943 (Fed. Cir. 2008) (en banc).
121 *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).
122 Kesselheim AS. The Supreme Court, process patents, and medical innovation. *N Engl J Med* 2009; **361**: 2303–6.
123 *Classen Immunotherapies v. Biogen Idec* (Fed. Cir., 2008).
124 US Patent 5,723,283.
125 Art. 52 et seq. European Patent Convention (EPC).
126 Art. 63 EPC.
127 Art. 52(2) EPC.
128 T 1173/97 (IBM).
129 European Patent Office. Guidelines for examination in the European Patent Office. 2017. https://www.epo.org/law-practice/legal-texts/guidelines/archive/guidelines-2017.html (accessed May 2019).
130 European Patent Office. Unitary Patent and Unified Patent Court. https://www.epo.org/law-practice/unitary.html (accessed May 2019).
131 European Patent Office. Unitary Patent. https://www.epo.org/law-practice/unitary/unitary-patent.html (accessed May 2019).
132 The 25 member states are Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Romania, Slovakia, Slovenia, Sweden, and the UK.
133 European Patent Office. Unified Patent Court. https://www.epo.org/law-practice/unitary/upc.html (accessed May 2019).
134 Abbott F, Cottier T, Gurry F. *International Intellectual Property in an Integrated World Economy*, 4th edn. New York: Wolters Kluwer, 2019.

Part III

Sources of Data for Pharmacoepidemiology Research

Part IIIa

Spontaneous Reporting

10

Postmarketing Spontaneous Pharmacovigilance Reporting Systems

Gerald J. Dal Pan[1], Marie Lindquist[2], and Kate Gelperin[1]

[1] *Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA*
[2] *Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden*

Potential signals for adverse drug reactions (ADRs) or adverse drug effects most often arise from postmarketing spontaneous case reports, which are collated and analyzed by drug safety experts, evaluated as clinical case series, and considered for potential regulatory action. These efforts are not possible without input from dedicated health professionals and other concerned stakeholders. Adverse events (AEs) thought to be potentially drug related may be reported by a consumer or a health professional to a drug's manufacturer, or they may be reported directly to a health authority through programs such as MedWatch or EudraVigilance [1,2]. In addition, case reports and case series with valuable clinical details may be published in a peer-reviewed journal [3]. Concerned stakeholders – health professionals as well as consumers – are the source of the signals that can trigger hypothesis generation, hypothesis testing, and appropriate regulatory action when needed to protect the public from unnecessary risks or harms. At times, a drug causal association may seem clear due to strong temporal association between exposure to the product and onset of an adverse effect, or when there is confirmation of positive rechallenge (i.e., signs or symptoms resolve when exposure is stopped but recur when reintroduced). But more often, causality assessment is challenging (see Chapter 29), and well-designed pharmacoepidemiology or clinical studies are needed to assess the signal [4,5].

In the United States, the Food and Drug Administration (FDA) issues Drug Safety Communications (DSCs) to alert the public about emerging safety issues, such as investigations into potential safety signals that may alter the balance of therapeutic benefit and risk for a medical product [6]. Recently, the FDA launched a new web portal that enables the public to view summary charts and listings of deidentified cases from the FDA Adverse Event Reporting System (FAERS), a compilation of all postmarketing adverse event (AE) reports received by the FDA [7].

In recent years, the term "pharmacovigilance" has become widely used to denote postmarketing

Note: The views expressed in this chapter are those of the authors, and not necessarily those of the US Food and Drug Administration or the US government.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

safety activities, and is defined by the World Health Organization (WHO) as "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other possible drug-related problems" [8].

Monitoring and understanding the safety of drug and therapeutic biologic products is a process that proceeds throughout the product's life cycle, spanning the period prior to first administration to humans through the entire marketing life of the product. Throughout the product life cycle, astute clinical observations made at the point of care constitute an important source of information. While new technologies have enabled more thorough knowledge of a drug's actions, and computerized databases have enabled large-scale, population-based analyses of drug safety investigations, these advancements are adjuncts to, and not substitutes for, careful, well-thought-out clinical observations.

Preapproval drug safety assessment includes animal toxicology and pharmacologic studies, first in humans studies (Phase I), proof-of-principle studies for the disease or condition under study (Phase II), and confirmatory studies of safety and efficacy (Phase III). In each of these stages of drug development, important drug safety information is obtained.

In the preapproval review process, regulatory authorities review these safety data, along with data on the product's efficacy, to determine if the anticipated benefits of the drug are likely to outweigh any risks with its intended use. In the US, as part of the approval process, the FDA reviews the professional labeling (package insert), to ensure that the product's uses and risks are explained adequately.

Although the preapproval testing of a drug is typically rigorous, and the review of the data is thorough, there are still inevitable uncertainties about the complete safety profile of a drug when it is brought to market. Several factors contribute to these uncertainties. First, the number of patients treated with the drug prior to approval is limited, generally from several hundred to a few thousand. Second, patients in clinical trials tend to be carefully selected for inclusion in these trials, and are thus more clinically homogeneous than patients treated in the course of clinical practice once a drug is marketed. Compared to patients in clinical trials, patients treated in clinical practice may have a broader range of co-morbidities, take a wider variety of concomitant medications, and have a wider clinical severity spectrum of the underlying disease being treated. Third, additional populations of patients, such as children or older adults, who may not have been studied in large numbers in premarketing clinical trials, may be treated with the product once it is marketed. In addition, marketed drug products are often used for diseases or conditions for which they are not indicated, or at doses outside the approved range. Because of this "off-label use," patients treated in clinical practice are more diverse than those treated in clinical trials. For these reasons, a postmarketing drug pharmacovigilance reporting system is necessary.

Description

Adverse Events and Adverse Drug Reactions

A key concept in pharmacovigilance is the distinction between the closely related, but nonetheless distinct, concepts of *adverse event* and *adverse drug reaction*. The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) E2D guideline on Post-Approval Safety Data Management: Definitions and Standards for Expedited Reporting defines an adverse event as follows [9]:

> An adverse event (AE) is any untoward medical occurrence in a patient administered a medicinal product and which does not necessarily have to have a causal relationship with this treatment. An adverse event can

> therefore be any unfavorable and unintended sign (for example, an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to this medicinal product.

The same guideline describes an adverse drug reaction as follows:

> All noxious and unintended responses to a medicinal product related to any dose should be considered adverse drug reactions.
>
> The phrase "responses to a medicinal product" means that a causal relationship between a medicinal product and an adverse event is at least a possibility.
>
> A reaction, in contrast to an event, is characterized by the fact that a causal relationship between the drug and the occurrence is suspected. If an event is spontaneously reported, even if the relationship is unknown or unstated, it meets the definition of an adverse drug reaction [9].

The principal difference between an adverse event and an adverse drug reaction is that a causal relationship is suspected for the latter, but is not required for the former. In this framework, adverse drug reactions are a subset of adverse events. In some countries, postmarketing pharmacovigilance reporting systems are focused on adverse drug reactions, while in others data on adverse events are collected. In the United States, for example, the scope of reporting requirements is "[a]ny adverse event associated with the use of a drug in humans, whether or not considered drug related …" [10].

While many of the principles discussed in this chapter apply equally to adverse events and adverse drug reactions, it is important to understand the distinction between these two concepts. Specifically, some databases may contain only adverse drug reactions, while others may contain adverse events. These databases may behave differently when used for data mining. However, because many of the principles of drug safety surveillance apply to both adverse events and adverse drug reactions, we will use the term "AE/ADR" to refer to these two terms collectively in this chapter, for convenience. When needed, we will use the individual terms if a distinction between the two is required. Although the medical literature may sometimes erroneously use these terms interchangeably, there has been increasing attention to the distinction [11].

The Concept of Spontaneous AE/ADR Reporting

A core aspect of pharmacovigilance is the voluntary reporting of AEs/ADRs either directly to established national or regional centers, or alternatively to pharmaceutical manufacturers, who in turn are obligated to report pertinent information to regulators. National reporting systems are typically run by regulatory agencies (e.g., the US FDA runs the MedWatch program) [1] or by centers designated by the health ministry or the drug regulatory authority. In a few countries, the national pharmacovigilance center is run by a university or other scientific body. In the United States for example, AEs/ADRs in individual patients are generally identified at the point of care. Patients, physicians, nurses, pharmacists, or anyone else who suspects that there may be an association between an AE/ADR and a drug or therapeutic biologic product are encouraged to, but are generally not required to, report the case to either the manufacturer or the FDA.

This system of AE/ADR reporting is often referred to as a spontaneous reporting system; "spontaneous" because the person who initially reports the AE/ADR to either the reporting center or the manufacturer chooses what events to report. Sometimes, spontaneous reporting systems are also labeled as "passive," based on the argument that the reporting

center or manufacturer passively receives this information rather than actively seeking it out. However, this term does not do justice to the proactive way in which many pharmacovigilance centers seek to operate, even if resource constraints often limit the ability to interact adequately with reporters. Moreover, "spontaneous reporting" does not fit well with the reporting situation of today, when most countries have introduced or enacted legislation which mandates reporting from pharmaceutical companies. Reporting may also include canvassed or stimulated reporting of suspected reactions of particular interest.

Underlying the concept of a spontaneous postmarketing AE/ADR pharmacovigilance reporting system is the notion that clinical observations made at the point of care are often valuable pieces of information in further refining the knowledge of a drug's safety profile. This is an important, though frequently underemphasized, idea.

First, after approval, when formal study often ends and marketing of the medicine begins, there is often no further systematic way to continue the study of a medicine's safety, or even to generate drug safety hypotheses. While scientific advances and access to new data sources (e.g., electronic healthcare records) may provide some opportunity to monitor the safety of a marketed medicine, these alternative approaches to safety signal detection remain unproven. Such sophisticated methods are not widely used in many regions, and when used, may cover a limited number of drugs and outcomes. In contrast, existing pharmacovigilance reporting systems apply to all marketed medicines and are relevant to most drug safety issues of interest.

Second, when healthcare professionals, patients, and consumers want to make notification of a potentially adverse effect of a medication, it is useful for this information to be systematically organized, stored, and analyzed. A reporting system fills this need. If such information were not systematically collected, potentially valuable data about medicines would be lost.

Third, this system implies an important role for healthcare professionals in postmarketing safety assessment. Although the practices and systems for healthcare professionals to report AEs/ADRs vary from region to region, the quality of reports is always dependent on the details provided by these professionals.

Spontaneous Reports and Solicited Reports

Another key concept in understanding the contents of a pharmacovigilance database is the distinction between a "spontaneous report" and a "solicited report." While many pharmacovigilance databases are often referred to as "spontaneous report databases," the reports in them are often a mix of spontaneous and solicited reports, as well as reports from other sources. The differences between these two types of reports can explain the quantity and quality of reports in a pharmacovigilance database, and often can explain important distinctions between pharmacovigilance databases.

The ICH E2D guideline on Post-Approval Safety Data Management: Definitions and Standards for Expedited Reporting defines a spontaneous report as follows [9]:

> A spontaneous report is an unsolicited communication by a healthcare professional or consumer to a company, regulatory authority or other organization (e.g. WHO, Regional Center, Poison Control Center) that describes one or more adverse drug reactions in a patient who was given one or more medicinal products and that does not derive from a study or any organized data collection scheme. Stimulated reporting can occur in certain situations, such as notification by a "Dear Healthcare Professional" letter, publication in the press, or questioning of healthcare professionals by company representatives. These reports should be considered spontaneous. Consumer adverse reaction reports should be handled as spontaneous reports irrespective of any subsequent "medical confirmation".

> Regulatory Authorities might require medical confirmation for the purpose of expedited reporting. Emphasis should be placed on the quality of the report and not on its source. Even if reports received from consumers do not qualify for regulatory reporting, the cases should be retained.

Several features of this definition are worth noting. First, by requiring that the report be directed to a pharmaceutical company, regulatory authority, or other organization responsible for surveillance of the adverse effects of medicines, the definition implies, but does not explicitly state, that the reporter specifically intended to report a suspected adverse drug reaction. However, in current practice, most pharmacovigilance reporting systems do not consider the reporter's intent when determining if a report is a spontaneous report. For example, a patient who has been on a chronic daily medicine for hypercholesterolemia for many years may contact that medicine's manufacturer to ask if there are any known drug interactions between that medicine and another product, such as an antiinflammatory agent that may have been recently prescribed for a sprained ankle after a sports injury. The intent of the call was to seek information, not to report a suspected adverse drug reaction. Nonetheless, this report meets the definition of a spontaneous report, at least in those systems in which adverse events, and not only adverse drug reactions, are collected. The consideration of stimulated reports as spontaneous reports is consistent with this logic.

Second, and importantly for pharmacovigilance systems that require reporting of adverse events and not only adverse drug reactions, the definition does not require a causality assessment. For the purposes of meeting adverse event reporting requirements, ICH E2D notes that "spontaneous reports associated with approved drugs imply a suspected causal relationship" [9]. It is important to note that this implied causal relationship is for the purposes of regulatory reporting, and need not represent a scientific or medical conclusion.

Third, the requirement that the report not derive from a study or an organized data collection scheme necessitates existence of another category of report to describe adverse events occurring in clinical trials, other studies, and certain organized programs sponsored by pharmaceutical companies that may collect data on adverse events. CIOMS V (Council for International Organizations of Medical Sciences) recognized the need to describe those adverse event reports derived not only from formal clinical trials or other studies, but also from the increasing number of company-sponsored programs, such as marketing programs and patient-support programs, that promote interaction between the company and patients – and thus the chance for companies to learn about adverse events [12]. CIOMS V proposed the idea of "solicited" reports, which was formalized in ICH E2D as follows [9]:

> Solicited reports are those derived from organized data collection systems, which include clinical trials, registries, post-approval named patient use programs, other patient support and disease management programs, surveys of patients or healthcare providers, or information gathering on efficacy or patient compliance. Adverse event reports obtained from any of these should not be considered spontaneous.
>
> For the purposes of safety reporting, solicited reports should be classified as study reports, and therefore should have an appropriate causality assessment by a healthcare professional or an MAH. Further guidance on study-related issues, such as managing blinded therapy cases, can be found in the ICH E2A guideline.

Unlike its recommendation for spontaneous reports, ICH E2D recommends that solicited reports be subjected to a causality assessment

for regulatory reporting; in general, only those serious adverse events deemed causally related are to be reported. ICH E2D also notes other types of adverse event reports, such as reports from the medical literature, reports from the internet, reports a company obtains in accordance with contractual relationships with another company, and reports a company receives from a regulatory authority. The first two types are considered unsolicited (though not spontaneous) while the latter two are considered solicited [9].

The CIOMS [12] and ICH E2D [9] efforts were initiated to provide a framework for regulatory reporting of adverse events, and many jurisdictions have incorporated the ICH E2D principles in their directives, regulations, guidelines, and guidance documents. While a discussion of regulatory reporting requirements is beyond the scope of this chapter, an understanding of the distinction between spontaneous reports and solicited reports is important because it is essential to understanding the contents of a pharmacovigilance database. For example, if a pharmaceutical company contributes the majority of reports to a particular pharmacovigilance database, that database can be expected to have more solicited reports, as well as spontaneous reports that result from interactions between the company's sales force and healthcare professionals, than a pharmacovigilance database whose reports are derived mainly from hospital-based pharmacovigilance centers. The distinction may also be important in the comparison of adverse event reports between two products in the same pharmacovigilance database. For example, in a single pharmacovigilance database, there may be more adverse event reports for a product that is actively promoted and marketed (and thus has a large sales force and one or more patient-focused marketing programs) than a product for which such an extensive marketing program is not in place.

Overview of Pharmacovigilance Reporting Systems

The goal of a postmarketing, or postapproval, safety program is to identify drug-related AEs or ADRs that were not identified prior to approval, to refine knowledge of the known adverse effects of a drug, and to understand better the conditions under which the safe use of a drug can be assured.

The scope of pharmacovigilance is broad. The core activity is usually the identification of previously unrecognized AEs/ADRs with use of the drug. However, it is not sufficient simply to note that use of a drug can lead to an AE/ADR. Rather, an investigation into not only the potential causal role of the drug in the development of the AE/ADR, but also the conditions leading to the occurrence of the AE/ADR in one person or population and not in others must be the focus of any postmarketing drug safety effort. Factors such as dose–response relationships, drug–drug interactions, drug–disease interactions, drug–food interactions, and the possibility of medication errors must be carefully considered.

A full understanding of the factors that can lead to an AE/ADR may yield ideas for effective interventions to minimize the severity or occurrence of the AE/ADR, and thus enhance the safe use of the drug. For this reason, the approach to detecting and understanding clinically important AEs/ADRs in the postmarketing period must be as comprehensive as possible.

The identification of a new safety issue with a medicinal product often begins with a single observation. Such observations may arise from animal studies, chemical studies and assays, or observations of human experience with the medicine. In the postmarketing period, such observations are usually clinical observations, often made at the point of care in the course of clinical practice. A practitioner or patient notes the development of symptoms or signs that were not present, or were present in less severe form, prior to the patient's using the medicine.

If this sign or symptom is not listed in the product's approved labeling, patients and healthcare professionals may not think to attribute it to the medicine. If further evaluation reveals a clinically significant process (e.g., liver injury, rhabdomyolysis, agranulocytosis), it is important to keep in mind the possibility of a side effect due to a medication in the differential diagnosis of the event. If a medication side effect is not included in the differential diagnosis, a potential association between a medicine and previously unrecognized side effect will not be made, and the patient may not be treated appropriately. If, on the other hand, the practitioner believes the medicine played a role in the development of the new clinical findings, he or she can forward relevant clinical information to either the medicine's manufacturer or to a drug regulatory authority, such as the FDA in the United States or other national or regional authorities, as appropriate.

In the postmarketing period, the investigation of AEs/ADRs is a multidisciplinary effort. The analysis of a complex AE/ADR can involve the fields of medicine, pharmacology, epidemiology, statistics, pharmacy, toxicology, and others. There are several methods of clinical postmarketing safety assessment. These include the review of case reports and case series from spontaneous reporting systems, a wide variety of types of observational epidemiologic studies, and clinical trials. This chapter will focus on spontaneous pharmacovigilance reporting systems. No one method is *a priori* better than another in all settings. Rather, the choice of methods depends on the particular safety question to be answered.

Spontaneous AE/ADR reports have at times served as a necessary and sufficient basis for regulatory actions including product withdrawals. For instance, in August 2001 the manufacturer of cerivastatin withdrew the drug from marketing based on "a markedly increased reporting rate of fatal rhabdomyolysis" compared to the other drugs in the statin class [13]. Additional confirmation of the unacceptably high risk of rhabdomyolysis with cerivastatin was eventually available three years later when results of a well-designed epidemiologic study were published [14]. Clearly, that time frame would have been far too long to delay decisive action, which in retrospect was soundly based on the signal from spontaneous reports. The timely detection of this signal would not have happened without the efforts of the point-of-care clinicians who took the time to report rhabdomyolysis when it occurred in their patients. Some drug safety experts have argued that decisive action could have been taken even earlier based on clinical trial data with a higher unapproved dose of cerivastatin, coupled with early postmarketing experience [15].

Patient Reports and Healthcare Professional Reports

Spontaneous adverse event reports, by their nature, originate at the point of care. While some pharmacovigilance systems were once restricted only to reports from healthcare professionals, there has been growing recognition of the importance of reports from patients, and many systems now accept patient-generated reports. For example, Italy, Denmark, the Netherlands, and Sweden have accepted patient reports since the early 2000s, while Australia has accepted them since 1964 [16]. The United States, which has accepted adverse event reports from consumers since 1969, developed the MedWatch program [17] in 1993 to facilitate adverse event reporting from both patients and healthcare professionals. Nonetheless, as recently as 2012, some countries with highly developed regulatory systems were not actively collecting patient reports. Of 50 countries with developed drug regulatory systems surveyed in 2013, 44 had direct patient reporting systems, 17 of which were started in 2012 or 2013 [16].

There is no internationally recognized definition of a "patient report." ICH E2D defines a "consumer" as "a person who is not a healthcare professional such as a patient, lawyer, friend, or

relative of a patient" and notes that "consumers" can submit adverse event reports [9]. Of note, in 2016 approximately 1.7 million reports were entered into the US FAERS; consumers were the source of about half (845 355) of these reports, the majority of which were submitted via pharmaceutical companies [7].

Despite initial skepticism about the value of patient reports [18], there is growing evidence that they are valuable because they often contain more detail than reports generated by healthcare professionals and can serve to complement those reports [19]. A study of the United Kingdom's Yellow Card system, which allows for AE reporting by both healthcare practitioners and patients, found that reports generated by patients had a higher median number of suspected adverse drug reactions per report compared to those generated by healthcare professionals, had a higher median word count, had more detailed information about symptoms, and more description of the emotional and social impact of the adverse event [20]. A study comparing adverse event reports submitted by patients and those submitted by healthcare professionals to the Dutch National Pharmacovigilance Center Lareb found that reports from patients were comparable to those from health professionals for the purpose of causality analysis [21]. Similarly, a recent UMC-Lareb collaboration assessed the contribution of patient reports to global signal detection in VigiBase, and concluded that patient reports provide unique information valuable in signal assessment, and recommended their inclusion in signal detection processes [22].

In addition to their value in describing adverse drug reactions, patient reports contribute to signal detection. A study of signals sent from the Dutch National Pharmacovigilance Center Lareb to the Dutch Medicines Evaluation Board found that the number of patient reports that contributed to a signal increased from zero in 2003 to 31 in 2008, and that the proportion of patient reports contributing to signal generation equaled their proportion in the database [23]. In one pharmaceutical company's AE database, signals were detected earlier when patient reports were included, compared to when only reports from healthcare providers were included [24]. Experience in the UK Yellow Card system suggests that, when analyzed separately from healthcare professional reports, patient reports may generate additional signals based on disproportionality [25].

Given the potential importance of patient reporting, efforts have been made to both encourage and simplify it. In 2012, the World Health Organization published a guide for countries to use in setting up a reporting system for the general public [26], which recommends that a patient reporting system ideally be established in the setting of an existing spontaneous reporting system. The reporting form for patients may be a dedicated patient-reporting form or the same form used by health professionals, but it should be understandable by a layperson. Education of the public on the importance of patient reporting as well as training of pharmacovigilance staff in the assessment of patient reports are other elements of the WHO guideline.

At the national level, a law passed in the US in 2008 required pharmaceutical manufacturers to include the following statement in direct-to-consumer advertising: "You are encouraged to report negative side effects of prescription drugs to the FDA. Visit www.fda.gov/safety/medwatch, or call 1–800–FDA–1088." This statement, however, did little to increase patient reporting of adverse events. In a sample of 123 drugs, the average monthly increase following the implementation of this statement was 0.24 reports per drug [27]. More recently, in 2013, the FDA introduced a consumer-friendly adverse event reporting form [28], following the introduction of which patient reporting increased by 36% [29].

Because most AE/ADR reporting systems rely on healthcare professionals, patients, and consumers to submit reports voluntarily, it is

generally recognized that there is substantial underreporting of AEs/ADRs via current systems. Two survey-based studies conducted in the US in the 1980s, one in Maryland [30] and the other in Rhode Island [31], examined physician reporting to the FDA and concluded that fewer than 10% of AEs/ADRs were reported to the FDA. These studies were conducted prior to the development of the current MedWatch program [1] in 1993, and do not consider the contribution of reporting from sources other than physicians.

Calculating the proportion of adverse event reports that a reporting system receives requires that the true number of AEs/ADRs in the population be known. For most AEs/ADRs, this number is not known or readily available. In some cases, however, data are available that allow an estimate of the extent of reporting to be calculated. For example, the extent of reporting to the FDA of cases of hospitalized rhabdomyolysis associated with statin use was estimated [32] using a projected estimate of the number of such cases in the US and comparing it to the number of reports of statin-associated hospitalized rhabdomyolysis in the FAERS, a database that houses the FDA's postmarketing adverse event reports. The projected national estimate was obtained by using incidence rates from a population-based cohort study, and applying those rates to national estimates of statin use. Across four statins (atorvastatin, cerivastatin, pravastatin, and simvastatin), the estimated overall extent of AE reporting was 17.7%. For individual statins, the estimated extent of reporting ranged from 5.0% (atorvastatin) to 31.2% (cerivastatin). Further analysis revealed that the high proportion of reporting of cerivastatin cases was driven by reports received after the dissemination of a Dear Healthcare Professional letter notifying physicians of the risks of cerivastatin-associated rhabdomyolysis. The estimated extent of reporting was 14.8% before the letter and rose to 35.0% after. It is important to note that the results of this study apply only to reporting cases of statin-associated rhabdomyolysis. The extent of reporting for different drug-event pairs will be different, and cannot be estimated from the results of this study.

Once reports are received by national pharmacovigilance centers, they are entered into AE/ADR databases. These databases can then be inspected for drug safety signals, which form the basis of further study, necessary regulatory action, or both.

Report Characteristics

The individual case report is the fundamental unit of a postmarketing pharmacovigilance reporting system. The extent to which such a reporting system can address specific drug safety questions depends, in large part, on the characteristics and quality of the individual reports. Specific report formats differ across jurisdictions, though many countries and regions collect information compatible with the ICH E2B format. The updated electronic messaging standard ICH E2B (R3) [33] specifies both administrative and product identification information, as well as information on the case. The standard is designed to work with a variety of national and international systems and incorporates endorsement of standards by participating standards development organizations such as the International Standards Organization (ISO), Health Level Seven (HL7), European Committee for Standardization (CEN), and Clinical Data Interchange Standards Consortium (CDISC) to enable wider interoperability across the regulatory and healthcare communities. Although potentially comprehensive in scope, the format also allows for limited data to be submitted. The principal domains of case information in the ICH E2B standard include patient characteristics, reaction(s) or event(s), results of tests and procedures relevant to the investigation of the patient, drug(s) information, and a narrative case summary and further information.

Regardless of the specific formatting requirements across jurisdictions, there are some fundamental components of an individual safety report that are important for a thorough review.

Product identification, in as much detail as possible, is essential for an assessment of a case report. For pharmaceuticals, the identification of the active ingredient(s) is critical to product identification. However, other factors can also be important, depending on the specific safety question. For example, the formulation of the product can be important, as certain active ingredients may be present in a variety of formulations. Many opioid agents come in oral, injectable, and transdermal formulations. Because the pharmacokinetic and other pharmaceutical properties can differ across these formulations, information on the formulation is important in determining if there are formulation-specific effects, including those that may result from medication errors. Additionally, if the drug safety question involves the assessment of an AE/ADR related to a product quality defect, information on both manufacturer and lot/batch number can be very important, as product quality problems typically involve specific lots from an individual manufacturer.

Reports describing medication errors, or the potential for medication errors, ideally contain information on the product involved, the sequence of events leading up to the error, the work environment in which the error occurred, and the type of error that occurred [34].

Characteristics of a good-quality case report have been published [33,34]. As discussed below, these characteristics include adequate information on product use, patient characteristics, medical history, and concomitant treatments, and a description of the AE/ADR, including response to treatments and clinical outcome. Our experience, based on many years of reviewing case reports, is that while a substantial amount of useful clinical information can be written in a succinct narrative, most narratives are incomplete, many to the extent that they are uninterpretable. While follow-up with the reporter is sometimes feasible for drug safety analysts during case review, this has been the exception rather than the rule, often due to resource constraints. Incomplete and uninterpretable case reports limit the effectiveness of postmarket pharmacovigilance reporting systems. Attempts to improve the systems will need to address the problem of poor case report quality rather than merely increasing the number of reports. Unfortunately, it is not unusual for the FDA to receive potentially important spontaneous reports which cannot be evaluated because of missing key information. For instance, 13 (2%) of a total 675 reports of hypersensitivity AEs/ADRs associated with heparin administration during an investigation of tainted heparin were excluded from an analysis of AERS data because the reports were "not interpretable" [35].

Information on product use should include the start date(s), stop date(s), doses, frequency of use, and indication for use. Dosage information is important in exploring dose–event relationships. Duration of use is important for characterizing the time course of AEs/ADRs relative to initiation of product use. Indication for use is also an important piece of information, as many products are used for more than one indication (either on-label or off-label). Certain AEs/ADRs may be related to specific indications. Alternatively, concomitant medications and other factors related to one indication but not others may confound interpretation of the AE/ADR. For these reasons, indication for use is an important element of a case report.

Patient information should include age, gender, medical history, and concomitant medication usage. The presence of factors that could confound the relationship of the drug to the AE/ADR, especially elements of the medical history and concomitant medication usage, are critical to the interpretation of individual case safety reports.

A description of the AE/ADR that allows for independent medical assessment is critical. A simple listing of coded diagnostic and procedure

terms is generally insufficient for adequate assessment of the report. A narrative of the event that includes the temporal relationship of drug usage to the development of the AE/ADR, the clinical and diagnostic features, the clinical course, any measures instituted to treat the AE/ADR, the response to these measures, and the clinical outcome are all essential components of a high-quality case report. Results of laboratory tests, imaging, and pathology results facilitate an independent interpretation of the report. Information on dechallenge (the resolution of the AE/ADR when the medication is withdrawn) and rechallenge (the redevelopment of the AE/ADR when the drug is reintroduced), if available, can be invaluable.

Social Media

Social media are a range of computer-based technologies that allow the creation and sharing of information, ideas, photographs, and other messages via electronic communication. User-generated content is a defining feature of social media. This content can be made available to others via computer-based networks that connect one user with other users or groups to form social networks. Depending on privacy settings, which in some cases may be chosen by the user, the user's content may be widely available to other users or it may be restricted to only certain users or groups. Given the widespread use of the internet and, to a lesser degree, of social media for health-related topics, there is interest in whether social media can be a source of drug safety signals or otherwise shed light on adverse drug reactions [36]. Because social media posts describe individual experiences, they can, in theory, describe adverse reactions to medicines. The use of social media for pharmacovigilance presents both opportunities and challenges [37,38].

With an estimated 2.5 billion users of social media worldwide [39], including in parts of the world where formal pharmacovigilance programs are not highly developed, social media have the potential to be a source of patient- and consumer-generated information about adverse events [37]. The ability to tap into this potential source of information is especially relevant given the growing importance of, and attention to, patient- and consumer-generated reports in pharmacovigilance.

The challenges of identifying drug safety signals in social media include all those inherent in traditional spontaneous reporting systems (e.g., underreporting, duplicate reports, lack of relevant details, and stimulated reporting) as well as additional ones posed by the unique features of social media. These latter challenges include the general lack of structure of social media posts, the often informal nature of writing, the use of "street names" for established pharmaceuticals without corresponding use of standard brand names or active ingredient names, the use of slang or other informal language to describe symptoms or other medical concepts, and the diffuse audiences in social media. A further challenge is that while there are millions of social media posts, only a small percentage will concern adverse drug reactions. These challenges might be expected to pose more problems in large general social media sites than in smaller, health-related social media sites. For example, a study using the general social media site Twitter to identify adverse drug reactions found that of 10 822 tweets that mentioned a drug of interest, an adverse drug reaction, determined by expert annotation, was present in approximately 1200 [40]. By contrast, a similar study based in the health-related social media site DailyStrength found that approximately 24% of posts that mentioned a drug also mentioned an adverse drug reaction [41].

In traditional adverse event reporting systems, data are entered using a structured format, such as the ICH E2B standard, and drug and adverse event information are coded using standard dictionaries and terminologies. Importantly, data collection methods in these traditional systems

usually are designed to collect relevant information and have the final data structure in mind. Social media do not share these characteristics. Because most social media posts do not concern drugs or drug-related adverse events, techniques to identify adverse events in social media must first identify mentions of drugs, and then must further identify mentions of drug–adverse event pairs. Once these pairs are identified, further analyses can begin.

The use of social media for pharmacovigilance is an area of active research to address these challenges [38]. Current research focuses on the use of natural-language processing and other techniques, such as supervised machine learning, to identify drug-related adverse events in social media [42]. One of the biggest challenges in this regard is the distinction of drug mentions associated with an adverse event from those with no association to an adverse event, a task complicated by the fact that though the former group is the relevant one, it is usually notably smaller than the latter group.

Another area of research is determining the utility of social media in pharmacovigilance. In 2014, the Innovative Medicines Initiative (IMI), a public–private partnership between the European Union and the European Federation for Pharmaceutical Industries and Associations, launched WEB-RADR: Recognising Adverse Drug Reactions to develop new technical tools to facilitate the detection and analysis of potential adverse drug reactions in social media sites. It also aimed to develop a mobile phone app for the reporting of suspected ADRs to regulatory authorities in the European Union (in the context of traditional adverse event reporting). One of several planned outgrowths of these efforts is the establishment of a regulatory framework for social media mining for adverse drug reactions [43].

Preliminary results of IMI WEB-RADR, based largely on analyses of posts in Twitter, suggest that some medicines receive more attention in social media relative to their frequency in VigiBase, while others receive much less. Individual Twitter posts were deemed to be not valuable, perhaps due to Twitter's character length restrictions; however, combining information from multiple posts generated by the same user was not examined [44]. Preliminary recommendations from IMI WEB-RADR for a regulatory framework note that data from social media should be treated as a "secondary use of data," the use of social media for signal detection and validation should be optional, reporting of individual case safety reports of adverse drug reactions from social media sites should not be required, and follow-up with social media users should not be required. Rather, drug manufacturers should include insights gained from social media regarding the safety of their products in the product's periodic safety update report or risk management plan [45]. In conclusion, more work is needed to refine the methods of extracting and analyzing data from social media for detection of adverse drug reactions.

National Pharmacovigilance Systems

The organization of postmarketing safety reporting systems and national pharmacovigilance systems varies around the world. The fundamental feature is that health professionals, and in some cases patients or consumers, are encouraged to send reports of AEs/ADRs to one or more specified locations. These locations can be the drug regulatory authority, an academic or hospital-based pharmacovigilance center (often working with or on behalf of a drug regulatory authority), or the drug manufacturer. The roles of these institutions vary from country to country, and depend greatly on the regulatory and national drug monitoring system in the country.

In low- and middle-income countries, with varying regulatory infrastructure, the focus in pharmacovigilance has been different from that in the more affluent parts of the world. Reports can result from counterfeit and substandard

drugs, known ADRs and drug interactions of concern to reporters, and ADRs resulting from medical error. In some countries, responding to queries about adverse reaction incidence, diagnosis, and management is a major part of the work of pharmacovigilance centers. In developing countries, there are often deficiencies in access to up-to-date information on drug safety that need remedying. On the other hand, large donations of new drugs to combat the endemic scourges of malaria, HIV/AIDS, tuberculosis, infestations, and other diseases, along with vaccines, have led to the high priority of monitoring their use for both safety and efficacy.

However, in many low- and middle-income countries there is currently not enough capacity for effective drug safety monitoring, and the improved access to new medicines adds additional strain on already overburdened or nonexistent pharmacovigilance systems. A survey from 2010 of pharmacovigilance systems in low- and middle-income countries found that seven of 55 responding countries indicated that they had no designated system in place, and fewer than half of the respondents had a budget for pharmacovigilance [46]. Consequently, lack of funding was mentioned as a hindrance to the development of pharmacovigilance, together with lack of training and a culture that does not promote AE/ADR reporting. Suggested key developments included training for health workers and pharmacovigilance program managers; active surveillance methods, sentinel sites and registries; and better collaboration between pharmacovigilance centers and public health programs, with a designated budget for pharmacovigilance included in the latter.

The WHO is now working together with major donor organizations to address the urgent need for capacity building in low- and middle-income countries. The strategy is focused on sustainable development, covering not only the implementation of reporting systems, technical support, and training of healthcare professionals, but also improvements in governance and infrastructure to support pharmacovigilance activities in the broader context of regulatory systems strengthening.

The perceived responsibility of healthcare professionals to report AEs/ADRs often varies around the world. Because the largest gaps in drug safety knowledge are believed to be for recently approved medicines, most countries emphasize the need to report AEs/ADRs, even less serious ones, for this group of medicines. For example, in the United Kingdom, recently approved drugs containing new active ingredients are marked in the British National Formulary with a black triangle [47], a symbol used to denote a product whose active ingredient has been newly licensed for use in the UK. In some cases, drug products meeting certain additional criteria are also marked with a black triangle, even if the active ingredient has been previously approved. The aim of the black triangle program is to prompt health professionals to report all suspected adverse reactions associated with the use of these products. In some countries, it is mandatory for physicians and dentists to report cases of suspected adverse drug reactions to the regulatory authority. Most countries, however, do not have such specific programs or requirements, but health professionals are encouraged to report and the national reporting centers provide general advice to health professionals on what events to report.

In a majority of countries, including countries in the ICH regions, other high-income countries, and 33 of 55 low- and middle-income countries responding to a 2008 survey [46], pharmaceutical companies that hold marketing authorizations are obligated to report AEs or ADRs to the regulatory authority. In some countries, the event is reportable only if an attribution of causality has been made. In other countries, the event is reportable even if no attribution has been made. For example, in the United States, pharmaceutical companies are required by law to submit spontaneous reports

of AEs/ADRs, regardless of attribution of causality, on an expedited basis if they are serious and unexpected. The AE/ADR is considered serious [1] when the patient outcome is death; life-threatening; hospitalization (initial or prolonged); disability; congenital anomaly; or requires intervention to prevent permanent impairment or damage. Periodic reporting of other types of AEs/ADRs, such as those considered serious and expected (labeled), or nonserious, is typically required as well. The periodicity of such aggregate reports is determined by the length of time the drug has been marketed, with increased frequency for newly approved drugs, and decreased frequency (e.g., annual) with older drugs.

While spontaneous reports of AEs/ADRs usually originate initially from the point of care, the more proximal source of reports coming into the national pharmacovigilance centers may vary from country to country. In countries outside the ICH regions, the majority of reports are received directly from physicians in hospital and in general practice. Cumulatively over the past 40 years, most reports in the ICH region have come from the point-of-care initial reporter via the pharmaceutical companies to the regulatory authority; however, in several EU countries (e.g., all the Nordic countries), reports coming directly from health professionals to the regulatory authority greatly exceed company reports during this period. The patterns are likely to change towards a higher proportion of company reports in those many countries where pharmaceutical companies are legally obliged to report AEs/ADRs. Some countries restrict reports to those received by physicians. Other countries accept reports from pharmacists, nurses, and patients. There is a current trend towards encouraging direct patient or consumer reporting, replacing the notion held by many in the past that such reports would not be a reliable and useful source of information.

In most countries, the national pharmacovigilance center is part of the drug regulatory authority; in some, the monitoring is carried out jointly by the drug regulatory authority/Ministry of Health and an independent institution. In Germany, the Federal Institute for Drugs and Medical Devices (BfArM) maintains a joint database for recording reported adverse drug reactions, together with the Drug Commission of the German Medical Profession. According to the professional code of conduct of physicians in Germany, all adverse drug reactions should be reported to the Drug Commission. In The Netherlands, the practical responsibility for postmarketing surveillance is shared between the Medicines Evaluation Board (MEB) and the Netherlands Pharmacovigilance Centre (Lareb). The MEB handles communications with market authorization holders; the role of Lareb is to process and analyze reports from health professionals and patients.

Decentralized drug monitoring systems exist both within and outside the ICH region. In France, the French Medicines Agency coordinates the network of 31 regional centers that are connected to major university hospitals. In the UK, there are four regional centers connected to university hospitals which have the special function of encouraging reporting in their regions. The reporting system in China involves 31 regional centers reporting to the National Center for Adverse Drug Reaction Monitoring in the China Food and Drug Administration (CFDA). In India, the Pharmacovigilance Programme of India has been in operation since 2010, with the Indian Pharmacopoeia Commission (IPC) running the National Coordinating Centre. The system is now operating nationwide, with 250 local monitoring centers in medical institutes.

National and International Postmarketing Safety Databases

Once submitted to the national drug safety monitoring program, individual case safety reports are stored in computerized postmarketing safety databases. Many national drug regulatory

authorities have databases which include suspected AE/ADR reports derived from a postmarketing reporting system, as well as reports from other sources, such as the published medical literature, and sometimes certain types of serious adverse events (e.g., those considered by a clinical investigator to be potentially caused by a study drug) from clinical trials. Examples of national reporting systems and databases include the Blue Card system (Australia), Canada Vigilance (Canada), the Canadian Adverse Events Following Immunization Surveillance System (CAEFISS) database (Canada), the French Pharmacovigilance Spontaneous Reporting System database (France), the Adverse Drug Reaction Information Management System of the Pharmaceutical and Medication Devices Agency, Ministry of Health, Labor, and Welfare (Japan), the Lareb database (Netherlands), the BiSi database (Sweden), the MHRA ADR database (United Kingdom), the FDA Adverse Event Reporting System (FAERS) database (United States), and the Vaccine Adverse Event Reporting System (VAERS) database (United States). In addition, there are two international reporting and database systems: EudraVigilance [2] in the European Union (run by the European Medicines Agency, EMA) and VigiBase [48] pooling data from the more than 120 member countries of the WHO International Drug Monitoring Programme (run by the Uppsala Monitoring Centre, UMC). VigiBase is also the system used as the national database by around 70 pharmacovigilance centers around the world; reports are stored directly in VigiBase but entered, managed, and analyzed remotely through an internet-based data management tool, VigiFlow.

To understand the results of an analysis of individual case reports from a postmarketing safety database, it is necessary to understand the unique features of the database, as each large postmarketing safety database differs from the others. It is necessary to understand if, and how, the data are coded. Many databases code drugs according to a local or national standard drug dictionary, while others use a standard international dictionary, such as WHODrug [49]. Similarly, many databases code individual AE/ADR reporter verbatim terms which describe the AE/ADR according to a standard medical dictionary, such as the Medical Dictionary for Regulatory Activities (MedDRA) [50]. In the ICH regions (Europe, Japan, and the United States), use of MedDRA is mandatory for coding of AEs/ADRs.

Beyond coding, several other features of the database are important to understand. First, does the database include only reports from postmarketing systems, or does it include reports from other sources, such as the medical literature or clinical trials? Second, does the database include reports only from health professionals, or does it also include reports from patients and consumers? Third, what is the range of medical products included in the database – drugs, biologics, blood, blood products, vaccines, dietary supplements? Fourth, does the database include reports from only one country or region, or does it include reports from regions outside the jurisdiction of the regulatory authority? Fifth, does the database include both "nonserious" and "serious" AEs/ADRs; if so, what proportion of the reports have been classified by the health authority (or other database manager) as serious? Sixth, does the database include all adverse events (i.e., events which may or may not be judged to be causally related to a medicine) or does it include only adverse drug reactions (i.e., events for which a likely causal relationship has been determined prior to entering the report into the database)? Seventh, how many individual case reports are in the database? Each of these factors is important in determining the utility of a particular database in answering a specific drug safety question.

Detecting Signals from a Postmarketing Safety Database

The impetus to use a postmarketing safety database to evaluate the potential relationship of an AE/ADR to a drug may come from various

sources. For example, postapproval animal studies may suggest that a certain AE/ADR may be associated with a drug. The finding that a particular member of a drug class is associated with a specific adverse effect may prompt a search for the same reaction in other members of the class. Publication of case reports or case series, or unanticipated safety findings from ongoing clinical trials can be important sources of new safety questions for a marketed product. These stimuli for more intensive review of AE/ADR reports are external to the database.

Identifying potential associations of AEs/ADRs to drugs using only information within the database involves the detection of signals. According to the WHO, a signal is "reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously" [51]. While there have been many definitions of a signal put forth over the years, the important underlying principle is that a signal is a hypothesis that calls for further work to be performed to evaluate that hypothesis. Signal detection is the act of looking for or identifying signals from any source.

In the setting of a relatively small number of reports, review of groups of reports or periodic summaries of reports has been a standard method of signal detection. For example, one could look at a list of all reports in which the outcome was "death" to see if this outcome was reported more frequently for some drugs than others. Summaries based on specific organ class toxicities could be reviewed to examine whether reports in one system organ class were proportionately more frequent for one drug than others. These methods depend on the ability of a drug safety specialist to recognize new or unusual patterns of case reports. While an astute specialist can identify signals using this method, this manual review is often neither practical nor reproducible for detecting signals from large postmarketing safety databases, some of which contain several million records.

In an effort to address this challenge, data mining techniques have been applied to pharmacovigilance AE/ADR databases. In broad terms, data mining refers to a process of analyzing data to find patterns. In the case of AE/ADR databases, most of these patterns would not be visible without the use of statistically based, computerized algorithms. A variety of specific algorithms have been applied to safety signal detection in AE/ADR databases (see Chapter 27) [52,53].

The fundamental feature of data mining techniques used to analyze adverse event databases is that each is based on finding "disproportionalities" in data; that is, the finding that a given AE/ADR is reported for a particular drug more often than would be expected based on the number of reports of that AE/ADR for all other drugs in the database. Several features of these methods are worth noting.

First, the methods are transparent. While the total number of reports for a drug varies over time (and may be highest in the first few years of reporting), this temporal trend will not necessarily alter the proportion of specific reactions for the drug. Thus, a given reaction may still be found to be disproportionately reported even as the total number of reports for the drug changes.

Second, these methods rely exclusively on reports within the database; no external data are needed. For this reason, understanding the characteristics of the database, as discussed above, is important. This feature has several consequences. Because the expected number of reports of a specific AE/ADR for a given drug (and thus the disproportionality of the drug–event pair) depend on the reports within the individual database, the degree of disproportionality for a given drug–event pair may vary from one database to the next. In the extreme, a given drug–event pair may have a strong signal of disproportionality in one database and no such signal in another. A second consequence is that as the background information for all drugs in the database changes, so does the expected number of reports of a specific AE/ADR for a given drug (and again the disproportionality of the drug–event pair).

Third, a signal of disproportionality is a measure of a statistical association within a collection of AE/ADR reports (rather than in a population), and it is not a measure of causality. In this regard, it is important to underscore that the use of data mining is for signal detection – that is, for hypothesis generation – and that further work is needed to evaluate the signal.

Fourth, the absence of a signal of disproportionality in a postmarketing safety database is not evidence that an important AE/ADR is not associated with a particular drug.

Data mining is sometimes done using a subset of an AE/ADR database; for example, a portion of the database limited to a specific class of drugs might be used to find relative differences in the frequencies of specific AEs/ADRs across the class [54]. Some of the data mining techniques used in pharmacovigilance have included the proportional reporting ratio, the reporting odds ratio, the Bayesian Confidence Propagation Neural Network (BCPNN), and the empirical Bayes method (also known as the Gamma Poisson Shrinker or the Multi-item Gamma Poisson Shrinker) [55]. As part of the IMI, a public-private partnership in Europe, the EMA established the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (IMI PROTECT) with a goal of conducting research to develop and test new tools for the benefit–risk assessment of marketed drugs. A range of signal detection algorithms were compared across seven spontaneous reporting databases, with no method found to be better than the others. Findings were inconsistent across databases. The choice of signaling criteria had a greater impact on signal detection performance than the choice of disproportionality methods [56,57].

Review of Case Reports

The review of individual case reports of AEs/ADRs is a complex process that has been described elsewhere [58,59,60]. It typically begins by identifying one or more case reports with the outcome of interest. Because the case reports that form a case series often come from disparate sources, it is usually necessary to develop a case definition. The case definition centers on the clinical characteristics of the event of interest, without regard to the causal role of the medicine whose relationship to the adverse event is being investigated. Once a case definition is established, each report is reviewed to determine if the event meets the case definition and if the report is to be included in the case series. Depending on the specific question(s) to be answered by the case series, other exclusion criteria may also apply. For example, one would always exclude a case in which the report suggests that the patient never took the medicine of interest. In other cases, one may restrict the case series to only certain formulations of the medicine (e.g., include case reports in which an intravenous formulation, but not an oral formulation, was used, if such exclusion is appropriate for the question at hand), or to certain age groups (e.g., limit the case series to only case reports describing the suspected adverse events in pediatric patients, if such exclusion is appropriate for the question at hand), or to certain indications for use (e.g., limit the case series to case reports in which the medicine was used for a certain off-label indication, if such exclusion is appropriate to the question at hand). Exclusion criteria for a case series must be carefully considered so that potentially relevant cases are not excluded, and all available information is fully assessed. In general, if the purpose of the case series is to examine the relationship between a medicine and a suspected AE/ADR that has not been previously associated with it, it is best to err on the side of inclusion to avoid missing clinically relevant, though incomplete, information about cases of interest.

Once the case series has been developed, it is next necessary to review each case report individually to determine whether there is a

plausible causal relationship between the medicine and the adverse event. At the level of the individual case report, it is often difficult to establish with certainty that the medicine caused the adverse event of interest (see Chapter 29) [61,62,63]. For example, if the AE/ADR of interest is already common in the population that takes the medication, establishing a causal role for the medicine in the development of the condition is generally not feasible using individual case reports or case series. For example, the incidence of Parkinson disease is much higher in persons over age 60 years than it is in persons below that age [64]. In this situation, review of a report describing a myocardial infarction in a 70-year-old patient on an antiparkinsonian agent will generally not be informative in determining if the agent played a causal role in the development of the myocardial infarction, as myocardial infarction occurs commonly in this age group. Similarly, review of a case report is not likely to shed light on the causal relationship between a medicine and an AE/ADR when the AE/ADR is a manifestation of the underlying illness which the medicine is treating. For example, review of case reports of worsening asthma in patients taking an antiasthma medication is not likely to be sufficient to establish a causal link between the worsening asthma and the medication.

Review of a case series to establish a causal relationship between a drug and an AE/ADR is most straightforward when the suspected AE/ADR: (1) is rare in the population when the medication is not used, (2) is not a manifestation of the underlying disease, (3) has a strong temporal association with drug administration, and (4) is biologically plausible as a drug reaction or is generally the result of a drug reaction based on other clinical experience. Examples of AEs/ADRs that often meet these criteria are acute hepatic failure, aplastic anemia, agranulocytosis, rhabdomyolysis, serious skin reactions such as Stevens–Johnson syndrome and toxic epidermal necrolysis, and certain arrhythmias, such as torsades de pointes.

The approach to assessing the causal role of a medicine in the development of an AE/ADR has evolved over recent decades. In general, the approach relies on a systematic review of each case report to ascertain the temporal relationship between drug intake and development of the adverse reaction, an assessment of any coexisting diseases or medications that could confound the relationship between the medicine and the AE/ADR, the clinical course after withdrawing the drug (dechallenge), and the clinical course after reintroduction of the drug (rechallenge), when applicable. Naranjo and colleagues described a method based on these general principles for estimating the likelihood that a drug caused an adverse clinical event [65]. The WHO has developed a qualitative scale for categorizing causality assessments [66].

In the development of a case series, once the individual cases are reviewed, it is important to integrate the findings across the cases to determine patterns that may point to a relationship between the drug and the AE/ADR. For example, does the AE/ADR appear at some doses but not at others? Does the AE/ADR appear after one or a few doses, or does it appear only after a more prolonged exposure? Is the spectrum of severity of the event homogeneous or heterogeneous? Are certain co-morbidities or concomitant medications more likely to be present in patients with the event? In the review of a case series, there are no prespecified answers to these questions that establish or exclude the possibility that the drug led to the AE/ADR. Rather, the characteristics of the individual cases, taken together with the patterns observed in the case series itself, can lead the analyst to determine if the medication has a reasonable possibility of causing the condition of interest.

Reporting Ratios

Because postmarketing safety reporting systems do not capture all cases of an event of interest, it is not possible to calculate an incidence rate for

a particular drug–event pair. However, analysis of AEs/ADRs based simply on numbers of reports, even after thorough analysis of these reports, does not in itself put these reports into the context of how widely a medicine is used.

To adjust for the extent of drug utilization in a population in the analysis of AE/ADR reports, a reporting ratio can be used. A reporting ratio is defined as the number of cases of a particular AE/ADR reported to a drug safety database during a specific time period divided by some measure of drug utilization in the same time period. Across drugs, the reporting ratios measure the relative frequency of the AE/ADR reports adjusting for differences in level of drug utilization. The numerator is derived from counts of AE/ADR reports associated with the drug of interest that are recorded in the postmarketing safety database during a specified time period. In the past, the denominator typically consisted of the number of dispensed prescriptions, used as a surrogate measure of drug exposure in the population over that same time period, and often estimated from proprietary drug utilization databases. The number of dispensed prescriptions was used because data on the number of unique individuals using the drug in a specified time period were generally not available.

More recently, such data have become available, and reporting ratios based on persons using the medication, and not prescriptions, are being calculated. In some cases, information is available on not only the number of persons receiving the drug or the number of prescriptions dispensed, but also on the duration of use. When such data are available, the denominator for the reporting ratio may be expressed in person-time. When using denominators based on person-time, it is important to be mindful of the assumptions of the person-time method, especially the assumption that events in the numerator occur uniformly over time. Because AEs/ADRs may not occur uniformly over time after a drug is started, this assumption does not always hold.

Because the reporting ratio (sometimes referred to as "reporting rate") is not a measure of incidence or prevalence, it must be interpreted cautiously. For AEs/ADRs that are rare in the general population (e.g., aplastic anemia), reporting ratios are sometimes compared to the background rate (incidence or prevalence) of that event in a defined population. In other situations, individual reporting ratios of a particular AE/ADR across different drugs used for a similar indication or within the same class are calculated and the magnitude of the differences in reporting ratios is compared. Interpretation of the comparison of reporting ratios across drugs must be undertaken with caution, since such comparisons are highly sensitive to variation in AE/ADR reporting and thus it is necessary to consider the differential underreporting of AEs in the postmarketing safety reporting system. The underlying assumption in estimating reporting ratios for comparison across a group of drug products is that each of the respective manufacturer's reporting practices for the drug of interest is similar over the reporting period. However, this assumption may not hold true in some cases, and a comparison of reporting ratios across drugs may not be valid.

Strengths

Signal Detection

The principal strength – and, arguably, the principal purpose – of a postmarketing safety reporting system is that it allows for signal detection, the further exploration of drug safety hypotheses, and appropriate regulatory decision making and action when necessary. As noted earlier in this chapter, signals can be detected by data mining methods, review of individual case reports, or assessment of case series. In many instances, further work is needed to determine with more certainty the relationship of the drug to the AE/ADR. The

capability for timely and effective signal detection is a key strength of a postmarketing pharmacovigilance reporting system.

Another key strength of a well-designed and effectively utilized postmarketing pharmacovigilance reporting system is that, in certain cases, the relationship of a drug to an AE/ADR can be established with sufficient confidence, usually by a case series, that necessary regulatory action can be taken. AEs/ADRs for which the relationship to a drug can be established with reasonable certainty are generally those that have a strong temporal association with drug administration, a low or near absent frequency in the underlying population, are not part of the underlying illness being treated, are generally the result of exposure to a drug or other toxin, and have no other likely explanation. Aplastic anemia, agranulocytosis, acute liver failure, rhabdomyolysis, certain arrhythmias such as torsades de pointes, and serious skin reactions such as Stevens–Johnson syndrome are examples of adverse events whose relationship to a drug can often be established by case series [67,68,69]. However, relative to all signals detected in a postmarketing safety reporting system, those about which a reasonably firm conclusion can be made on the basis of AE/ADR reports alone are few in number.

Opportunity for the Public to Report AEs/ADRs

Postmarketing safety reporting systems allow healthcare professionals to report suspected AEs/ADRs to national pharmacovigilance centers, drug regulatory authorities, and/or manufacturers. Such systems allow for direct engagement of healthcare professionals in the drug safety monitoring system. The advantage of this involvement is that it allows for careful clinical observations, made at the point of care, to inform drug safety surveillance. Clinicians can provide succinct but detailed accounts of relevant symptoms, signs, diagnostic test results, past medical history, concomitant medications, and clinical course of an AE/ADR, including information on dechallenge and rechallenge. Such a synthesis of clinical information is generally not available from automated data sources. For those AEs/ADRs that are serious, rare, and often the result of a medication exposure, the ability to obtain detailed information directly from the point of care is an essential feature of postmarketing pharmacovigilance reporting systems.

Postmarketing safety reporting systems also can accept reports from consumers and patients, though this practice is not a feature of all reporting systems. In the US, where consumers and patients can report either to the manufacturer or directly to the FDA, the percentage of reports in 2016 that originated from consumers was about 50% [7]. When consumer- and patient-generated reports do not contain sufficient medical detail for meaningful review, subsequent follow-up with health professionals may be possible in potentially important cases, so that more complete clinical information (e.g., hospital discharge summary) can be obtained.

Scope

The scope of a postmarketing safety reporting system is quite broad. The system can cover all medicines used in the population, and it can receive reports of AEs/ADRs occurring in any member of the population. Because it need not restrict the reports it receives, it can receive AE/ADR reports throughout a medicine's marketed life cycle. Thus, AEs/ADRs recognized late in a product's life cycle, such as those resulting from prolonged exposure to a medicine, can, in theory, be ascertained. In practice, such ascertainment is difficult to achieve, because healthcare professionals may be less likely to ascribe an AE/ADR not known to be associated with a medicine that has been marketed for several years. In addition, patients who take a medicine for several years may also receive other treatments during that

time, making it difficult to conclude that there is an association between the medicine and the AE/ADR.

Despite this broad scope, a postmarketing spontaneous reporting system can be relatively inexpensive. Most of these pharmacovigilance systems rely on voluntary reporting, and those who report AEs/ADRs are generally not paid. Thus, information collection is not expensive from the perspective of effective pharmacovigilance, given that the system has the capacity to handle all medicines and all outcomes. This is in contrast to other data used to study drug safety questions, such as data from clinical trials, registries, and electronic healthcare data, each of which is relatively expensive to operate.

Limitations

Quality of Reports

Perhaps the major potential limitation of a spontaneous postmarketing safety reporting system is that it depends quite heavily on the quality of individual reports. Although data mining and other informatics methods can detect signals using coded bioinformatics terms in safety databases, each individual case report must still be carefully reviewed by a clinical analyst to determine if there is a plausible relationship between the medicine and development of the AE/ADR. The quality of the report, as described earlier in this chapter, is critical for an informative and meaningful review. Report quality depends on the care, effort, and judgment of the person submitting the report, as well as the diligence of the person receiving and/or transmitting it to the health authority. Reports without sufficient information for an independent determination of the relationship between the medicine and the AE/ADR are problematic for drug safety surveillance. However, with successful follow-up, sometimes even such deficient reports can yield useful information.

Underreporting

Another well-recognized limitation of spontaneous postmarketing reporting systems is underreporting. Because most systems are voluntary, not all AEs/ADRs are reported. A consequence of underreporting is that population-based rates of AEs/ADRs cannot be calculated, because all such occurrences in the population are not reported and the extent of underreporting for any individual AE/ADR is not known. Reporting ratios, discussed earlier in this chapter, allow the reported number of AEs/ADRs to be put into the context of drug utilization, though this measure is not an incidence rate.

Nonuniform Temporal Trends in Reporting

Another limitation of spontaneous reporting systems is that temporal trends in the number of AE/ADR reports for a drug–event combination may not reflect actual population-based trends for that combination. This is because multiple factors can affect the number of AE/ADR reports received for a given drug–event pair.

First, the number of reports for a medicine is thought to peak in the second year after approval and decline thereafter, even though the drug may be used more widely. This phenomenon, known as the Weber effect, was originally described in relation to nonsteroidal antiinflammatory medicines [70]. A more recent analysis of reporting patterns for the angiotensin II receptor blocker class of medicines revealed no discernible trend when the number of reports over time was examined [71]. Specifically, this analysis did not confirm that the number of reports increased toward the end of the second year and declined thereafter. Rather, it indicated that additional factors, such as the approval of additional indications and modifications of the firms' reporting requirements, affected the total number of reports received. However, when the

number of reports in a year was adjusted for the number of prescriptions dispensed in that period, it was found that the adjusted number of reports was highest in the first years after approval and declined thereafter. The frequency of AE/ADR reports per estimated unit of drug utilization may not be constant over time, although a recent analysis of FAERS data for 62 drugs did not confirm a reporting pattern over time as described by Weber [72].

Second, publicity about an important new AE/ADR often gives rise to a large number of reports shortly after the publicity, with a decline in the number of reports shortly thereafter. This phenomenon is known as stimulated reporting and was observed, for example, in the reporting pattern of statin-induced hospitalized rhabdomyolysis after publicity of this risk. For these reasons, changes in the number of AE/ADR reports for a given drug–event pair cannot reliably be interpreted as a change in the population-based frequency of the AE/ADR.

Another limitation of a postmarketing reporting system is that it is usually not well suited to ascertaining the relationship of a medicine to an AE/ADR that is common in the treated population, especially if the condition is a manifestation of the underlying illness. In such cases, the combined effect of confounding of patient factors and indication makes causality assessment of individual cases difficult.

Finally, duplicate reports of the same AE/ADR may be received by drug manufacturers and health authorities and if undetected as duplicates, may be entered into the database as multiple occurrences of the same event. Algorithms have been developed and various methods can be used to identify such reports; nonetheless, this issue is a potential source of bias and limits the utility of data mining or other calculations which rely on "crude" case counts which have not been "deduplicated."

Particular Applications

Fingolimod

Fingolimod, a sphingosine-1-phosphate receptor modulator that reduces the number of lymphocytes in peripheral blood, is used to reduce the frequency of clinical exacerbations and delay the accumulation of physical disability in patients with relapsing forms of multiple sclerosis. When the product was initially approved in the US in September 2010, the product label noted a dose-dependent reduction in peripheral lymphocyte count of 20–30% of baseline values and warned of the risk of serious infections. Based on experience in premarketing clinical trials, the label described the specific risk of fatal herpetic infections in two patients who received a dose higher than the recommended dose for multiple sclerosis. The label also noted that while the overall rates of infections and serious infections were similar in fingolimod-treated and placebo-treated patients in clinical trials, bronchitis and pneumonia were more common in fingolimod-treated patients.

Approximately three years after approval in the US, a patient in Europe was reported to have developed progressive multifocal leukoencephalopathy (PML), a demyelinating central nervous system disease caused by the JC virus. PML is a rare disease; when it occurs, it is usually in the setting of immunosuppression. Because the patient in whom PML was reported had received prior immunosuppressants, the PML could not be conclusively linked to fingolimod. At the time this case was reported, approximately 71 000 patients worldwide had received fingolimod, according to the manufacturer.

Two years after the initial case report of PML, the US FDA announced that it had received two case reports of fingolimod-treated patients with no prior immunosuppressant treatment [73]. In the first case, a 49-year-old patient with a five-year history of multiple sclerosis was suspected to have PML when results of a routine magnetic

resonance imaging test (MRI) showed lesions, not present at the time fingolimod treatment was initiated, that were atypical for multiple sclerosis and more consistent with PML. The patient had prior treatment with interferon beta-1a and intermittent corticosteroids. Cerebrospinal fluid analysis was positive for JC virus DNA. Based on the MRI findings and cerebrospinal fluid analysis, a diagnosis of probable PML was made, in accordance with the diagnostic criteria of the American Academy of Neurology consensus statement [74].

The second case concerned a 54-year-old patient with a 14-year history of multiple sclerosis who had been taking fingolimod for 2.5 years when PML was diagnosed. The patient had previously been treated with interferon beta-1a for 11 years, and was then switched to treatment with fingolimod. At the time of the PML diagnosis, the patient had been receiving treatment with mesalazine for ulcerative colitis for four years. After 2.5 years of treatment with fingolimod, the patient developed walking instability, clumsiness, inattention, and somnolence. A brain MRI was suggestive of PML, and cerebrospinal fluid analysis was positive for JC virus DNA. Based on the symptoms, MRI findings, and cerebrospinal fluid analysis, a diagnosis of definite PML was made, in accordance with the American Academy of Neurology diagnostic criteria. On the basis of these two cases, the product label for fingolimod in the US was updated to include a warning for progressive multifocal leukoencephalopathy.

This example illustrates some important features of the analysis of individual case safety reports. First, individual reports can be used to establish a causal relationship between a drug and an adverse event when the adverse event is rare in the population. In this example, PML is extremely rare, and when it occurs, it is usually in the setting of immunosuppression due to treatment with certain medicines or certain malignancies. If PML occurred spontaneously (i.e., in the absence of these particular conditions) in the general population or in patients with multiple sclerosis, it would be quite difficult to establish a causal relationship between PML and fingolimod therapy. Second, detailed, though not necessarily lengthy, case reports are important for robust case analysis, and especially for causality assessment. The two case reports in this example included detailed information about prior and concomitant medications (none of which is known to cause PML). Without information on prior and concomitant medications, the uncertainty about exposure to immunosuppressants would have limited the conclusions that could be made from the case reports. Third, the inclusion of relevant clinical information used to establish the diagnosis of the adverse event allows reviewers of the case report to come to an independent conclusion about the diagnosis. In this case, the details of the diagnosis were applied to the published consensus-driven diagnostic criteria set forth by a professional society. While use of such criteria is ideal, formal, established diagnostic criteria are not available for all adverse outcomes of interest, and reviewers of individual case safety reports should establish their own criteria.

Finally, this example is unusual in that a causal relationship between a drug and a serious AE was established based on two individual case reports. Because establishing a diagnosis and evaluating causality are difficult with individual case safety reports, a higher number is usually needed.

Dabigatran

Dabigatran is an oral direct thrombin inhibitor approved in the US in October 2010 to reduce the risk of stroke and systemic embolism in patients with nonvalvular atrial fibrillation. It was the first oral anticoagulant approved for this indication since warfarin had been approved for a similar indication. In the clinical trial that supported dabigatran's approval, 6076 patients were randomized to dabigatran 150 mg twice daily, and 6022 were randomized to warfarin

treatment. The rates of major bleeding were 3.3 per 100 person-years in dabigatran-treated patients and 3.6 per 100 person-years in warfarin-treated patients (hazard ratio 0.93, 95% confidence interval (CI) 0.81–1.07). The corresponding rates for life-threatening bleeds were 1.5 per 100 person-years and 1.9 per 100 person-years in the dabigatran and warfarin groups, respectively (hazard ratio 0.80, 95% confidence interval 0.66–0.98). Gastrointestinal bleeding was more common in dabigatran-treated than in warfarin-treated patients (1.6% vs 1.1%, hazard ratio 1.5, 95% confidence interval 1.2–1.9). In the first 14 months after dabigatran's approval, the US FDA received 2347 case reports of bleeding with dabigatran (348 with a fatal outcome), compared to 647 case reports of bleeding with warfarin (46 with a fatal outcome) [75].

On its face, this disparity in report numbers between the two agents suggested that dabigatran might be responsible for more bleeding in actual practice, a finding that was contrary to the preapproval observations. These numbers also raised the possibility of increased mortality with dabigatran relative to warfarin. A population-based analysis using administrative claims data, however, found that the rates of bleeding associated with dabigatran use were no higher than those associated with warfarin use [76]. A subsequent study using data from the US Medicare system found that, in actual practice settings, dabigatran was associated with a lower risk of ischemic stroke, intracranial hemorrhage and death and a higher risk of major gastrointestinal bleeding, relative to warfarin – findings that were consistent with those of the preapproval clinical trial data [77].

This example illustrates some important limitations concerning use of aggregate spontaneous report data to estimate population-based risk or relative risk of an adverse event between two drugs. A comparison of raw numbers of adverse event reports generally cannot be used to estimate the relative frequency of the adverse event in a population between two drugs. Population-based rates of an adverse event generally cannot be estimated from spontaneous reporting data because of underreporting and lack of a reliable measure of population exposure. Importantly, there often is a differential extent of underreporting of adverse events across a product's marketed life. Studies of adverse event reporting patterns of angiotensin receptor blockers and antiepileptic drugs have shown that, after adjustment for drug utilization, there were more adverse event reports in the first year after approval compared to subsequent years [71,78].Thus, even though a newly approved drug may not be widely used, there may be more spontaneous adverse event reports received for the newer drug compared to an older, widely used drug, even if there is no true difference in risk between them.

Peginesatide

Peginesatide, a novel synthetic peptide which was considered to be an important breakthrough for treatment of anemia in patients with dialysis-dependent chronic kidney disease, was withdrawn from marketing in 2012, within months of becoming commercially available, when the manufacturer received an unexpected number of case reports of fatal anaphylaxis [79]. Subsequent analyses suggested that the root cause of the reactions may have been a preservative present in the commercially available multiple-use vials but not in the single-use formulation that had been used exclusively in clinical trials [80,81].

Anaphylaxis, whether mediated by immunologic or nonimmunologic mechanisms, is a rare, unpredictable adverse reaction that can occur within minutes of exposure to an offending agent. Such reactions can alter the benefit–risk balance of newly approved drugs or biologics. Because fatal anaphylaxis is a rare occurrence with a strong temporal relationship between a triggering exposure and the onset of severe

symptoms, it represents a good example of the type of adverse drug reaction for which postmarketing spontaneous reporting systems are the primary means of timely signal detection. Although the decision to withdraw peginesatide from marketing was relatively swift, the results of subsequent nonclinical analyses suggesting an unexpected root cause illustrate the role of multiple data streams [82].

Drugs to Treat Attention-Deficit/ Hyperactivity Disorder

In other examples of postmarketing pharmacovigilance issues, spontaneous reports have provided actionable information about the clinical spectrum of adverse drug effects that may not have been well recognized in more restrictive clinical trial settings. An FDA safety evaluator became aware of several spontaneous reports describing psychiatric adverse events in otherwise normal children who were being treated with an extended-release formulation of methylphenidate for attention-deficit/ hyperactivity disorder (ADHD), and presented her findings at a Pediatric Advisory Committee meeting in June 2005. Committee members expressed concern, and a comprehensive evaluation of psychiatric adverse effects with drug treatments of ADHD was undertaken with the full cooperation of the drug's manufacturers. The results of the analysis were presented at a subsequent Pediatric Advisory Committee meeting in March 2006 and were also later published in a peer-reviewed journal [83]. Data were analyzed from 49 randomized controlled clinical trials. Results showed a total of 11 psychosis/mania adverse events which occurred during 743 person-years of double-blind treatment with the drugs of interest, compared to no similar adverse events during 420 person-years of placebo exposure in the same trials. Analysis of postmarketing spontaneous data yielded a total of 865 unique reports of psychosis or mania-type adverse events associated with these drugs. These findings were the basis for a MedWatch Alert in 2007, and for the addition of new warnings and medication guides for all of the ADHD drug treatments which were studied [84].

Medication Errors

The detection of medication errors is now an established area of pharmacovigilance [85,86]. An analysis of the EudraVigilance database revealed that between 2002 and 2015, a total of 147 824 cases of medication errors had been reported, with the annual number of such reports increasing throughout that period. Between 2010 and 2015, case reports of medication errors accounted for 1–2% of all reports in EudraVigilance [87].

For the purposes of pharmacovigilance, there is not an internationally accepted definition of a medication error. In the US, the National Coordinating Council for Medication Error and Reporting defines a medication error as "any preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of the health care professional, patient, or consumer" [88]. In the European Union, the Good Practice Guidance defines a medication error as "an unintended failure in the drug treatment process that leads to, or has the potential to lead to harm to the patient." Regardless of the specific definition used, preventability is a key concept underlying the detection of medication errors. Careful analysis of case reports of medication errors can lead to changes in the design of a product, changes to its instructions for use, changes to carton or container labeling, or other changes aimed at reducing the frequency of preventable errors.

Because a medication error can occur at many points in the drug use process, attentive healthcare professionals, patients, and others can detect the error (such as a wrong drug or wrong dose) before it reaches the patient – a so-called "near miss." Thus, unlike case reports of adverse

events, which describe an adverse outcome associated with use of a medicine, a case report of a medication error describes a medication error along any point in the medication use process, whether or not the patient receives the medication or experiences harm. Analysis of case reports of medication errors focuses on the product(s) involved, the sequence of events leading up to the error, the work environment in which the error occurred, the personnel involved (healthcare professionals, patient, family, and others), the type(s) of errors that occurred, and other contributing factors. For these reasons, case reports of medication errors should include detailed information on these factors [34]. The Medical Dictionary for Regulatory Activities (MedDRA) [50] contains terminology for medication errors, which allows for both broad and narrow searches of reports of medication errors in pharmacovigilance databases.

For example, the antidepressant vortioxetine was approved in the US in September 2013 and carried the proprietary name Brintellix®. By June 2015, the US FDA had received 50 case reports describing medication errors in which the name Brintellix was mistaken for Brilinta®, the proprietary name of the antiplatelet agent ticagrelor. Review of the case reports indicated that the wrong medication was dispensed in at least 12 cases, though there were no reports of ingestion of the wrong medicine. The case reports also indicated that this medication error occurred both when prescribing the medication and when dispensing it. Based on analyses of the two proprietary names, the confusion was likely due to the drugs' having the same first three letters, being presented near each other in a computerized order entry system, lack of pharmacist familiarity with the recently approved medication Brintellix, and the drug names looking and sounding similar to each other.

Between the analysis of these cases and a subsequent regulatory action, the FDA received an additional five case reports of Brintellix being confused with Brilinta. In one case, the medical record of a patient undergoing a lung biopsy indicated that the patient was taking Brilinta but the medical staff confused it for Brintellix. Not aware that the patient was taking an antiplatelet agent at the time of the lung biopsy, the medical staff did not take the necessary precautions and the patient experienced bleeding and a collapsed lung. In May 2016, the FDA approved a change in the proprietary name from Brintellix to Trintellix [89].

Data Mining Signals

According to the UMC glossary of pharmacovigilance terms, a signal is "a hypothesis of a risk with a medicine, with various levels of evidence and arguments to support it" [8]. Signals are identified by UMC analysts from the WHO Global Individual Case Safety Report (ICSR) database (VigiBase) by applying a predefined triage algorithm (data mining). The disproportionality measure used by the UMC is the information component (IC), originally introduced through the BCPNN, which is a logarithmic measure of the disproportionality between the observed and expected reporting of a drug–event pair. A positive IC value means that a particular drug–event pair is reported more often than expected, based on all the reports in the database. Signals from VigiBase are reported quarterly in the Signal document, which is circulated in restricted fashion to national pharmacovigilance centers for the purpose of communicating the results of UMC evaluations of potential data mining signals from the WHO database. A recent analysis found that of 43 UMC signals disseminated between 2007 and 2010 for products with approved labeling, 15 (35%) were labeled, and eight of the labeled signals were subsequently updated after the signal communication, supporting the relevance of routine data mining [90].

Below is an example of a WHO program signal identified by data mining applied to the WHO Global Individual Case Safety Report Database, VigiBase.

Topiramate and Glaucoma

Topiramate was approved in the US in 1996 as an anticonvulsant drug [91]. In the second quarter of 2000, reports of topiramate and glaucoma in VigiBase reached the threshold of an "association" (i.e., the lower limit of a 95% Bayesian confidence interval for the IC exceeded zero). When potential signals are identified, the available information is reviewed by the UMC staff and an expert review panel. At the time, there were six cases reported to VigiBase. After review, a summary of the findings were circulated in the Signal document in April 2001 to all national pharmacovigilance centers in the WHO program. On September 26 the same year, the Market Authorization Holder issued a Dear Healthcare Professional letter warning about "an ocular syndrome that has occurred in patients receiving topiramate. This syndrome is characterized by acute myopia and secondary angle closure glaucoma." By August 17, there were 23 reported cases according to the company. The FDA issued a warning in the revised labeling on October 1, 2001 [91].

Signals from Developing Countries

At the annual meetings of the WHO program members, country representatives are invited to share problems of current interest in their countries. Below are two examples illustrating the kind of issues that have been investigated in developing countries, presented at the 2017 meeting in Uganda [92].

Blindness and Retinal Disorder Associated with Clomifene Citrate: Case Series Assessment

A case of retinal detachment with the use of clomifene citrate that caused irreversible blindness triggered an assessment by the Eritrean Pharmacovigilance Centre. A search of VigiBase identified 24 cases of blindness and retinal disorder. All cases were evaluated using Austin Bradford Hill considerations to assess the causal relation. In all cases, clomifene was reported as the sole suspected drug and in all but three cases, no concomitant drugs were reported. There were two cases of blindness in which the reaction abated with sequelae following withdrawal of clomifene. The conclusion was that the findings support a causal relationship and warrant further investigation to substantiate the signal [93].

Signal of Alpha-Chymotrypsin and Anaphylaxis

Alpha-chymotrypsin is a biological product commonly used in Vietnam for numerous conditions. The efficacy of the oral product was questioned because this product is a large molecule product and the potential for safety issues of the injectable form warranted an investigation. From the national spontaneous reporting database, significant signals related to hypersensitivity, including anaphylactic reactions (reporting odds ratio [ROR] 2.12; 95% CI 1.46–3.07). Since 2010, 249 reports were received nationwide, of which 65 cases were related to anaphylactic reactions, and this is approximately equal to all spontaneous reports related to alpha-chymotrypsin obtained from VigiBase. The National Centre sent an official letter to the Drug Administration of Vietnam, Ministry of Health to advocate a safety effectiveness revision for this product [92].

Deployment of Pharmacovigilance During Mass Drug Administration in Sierra Leone

The specific challenges for pharmacovigilance in a developing country during a public health emergency were illustrated by Wiltshire Johnson, Registrar and CEO of the Pharmacy Board of Sierra Leone during the May 2016 Uppsala Forum conference [94]. Pharmacovigilance during the Ebola crisis in 2014 meant not only the safety surveillance of experimental treatments, but also of products such as disinfectants, rubber gloves, and other equipment. Treatment of malaria, as well as pneumonia and diarrhea, became difficult due to the reluctance to seek medical help; many feared that the similarity of symptoms with Ebola would prevent them from returning home. The surveillance of the antimalarial mass drug administration

of over 5 million doses of artesunate-amodiaquine that needed to be rolled out during the peak of the Ebola outbreak stretched all capacity, but became a success story. In collaboration with the National Malaria Control Programme, the Pharmacy Board's National Pharmacovigilance Centre actively participated in real-time pharmacovigilance by going into the communities and searching for, identifying, and managing adverse events. The data analysis led to changing first-line treatment of malaria from artesunate-amodiaquine to artemether-lumefantrine [94].

The Future

Spontaneous AE/ADR reporting is an important component of drug safety surveillance. The widespread availability of electronic healthcare data may, at first, seem to undermine the importance of AE/ADR reporting. This is not likely to be the case. Because careful observation at the point of care is an essential component of pharmacovigilance, electronic systems may be able to facilitate AE/ADR reporting in the future but will not replace it. It is technologically and administratively feasible for carefully designed systems to allow clinicians to report AEs/ADRs directly from electronic medical record systems. If designed properly, these systems could allow for the accurate, complete, and efficient inclusion of laboratory, radiologic, and other diagnostic test results, information which is often incomplete in current AE/ADR reports. The challenge of such a system will be to encourage reporters to routinely provide a clinically meaningful narrative that explains concisely the clinical course of the AE/ADR and its relationship to medication usage.

There is also interest in using modern informatics techniques to facilitate review of adverse event reports, especially in large AE databases. For example, the use of natural language processing techniques is being explored to determine if they can identify individual case safety reports that warrant further evaluation, or individual case reports that suggest a causal association between a medicine and an adverse event. Postmarketing safety reporting systems depend on the involvement of healthcare professionals and, in some areas, consumers and patients as well, for high-quality AE/ADR reports.

As new medicines become available, it will be increasingly necessary to monitor postmarketing safety. Postmarketing safety reporting systems will continue to be the cornerstone of this effort, because of their unique advantages. As social media, active surveillance, and the use of large healthcare databases begin to play a role in drug safety surveillance, demonstrate their utility, and realize their potential, they could become valuable adjuncts to existing pharmacovigilance reporting systems worldwide.

References

1 FDA. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. www.fda.gov/Safety/MedWatch/default.htm (accessed April 8, 2019).

2 EMA. EudraVigilance. www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000679.jsp (accessed April 8, 2019).

3 Russmann S. Case reports of suspected adverse drug reactions: case reports generate signals efficiently. *BMJ (Clinical Research Ed)* 2006; **332**(7539): 488.

4 Naranjo CA, Busto U, Sellers EM, *et al*. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Therapeut* 1981; **30**(2): 239–45.

5 Ralph Edwards I. Causality assessment in pharmacovigilance: still a challenge. *Drug Saf* 2017; **40**(5): 365–72.

6 FDA. Drug Safety Communications. https://www.fda.gov/Drugs/DrugSafety/ucm199082.htm (accessed April 8, 2019).

7 FDA Adverse Events Reporting System (FAERS) Public Dashboard. https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070093.htm.

8 UMC/WHO. Glossary of Pharmacovigilance Terms. www.who-umc.org/global-pharmacovigilance/global-pharmacovigilance/glossary/ (accessed April 8, 2019).

9 ICH. Post-Approval Safety Data Management: Definitions and Standards for Expedited Reporting E2D. ICH Harmonised Tripartite Guideline. www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2D/Step4/E2D_Guideline.pdf (accessed April 8, 2019).

10 US Food and Drug Administration. 21 CFR 314.80. www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.80 (accessed April 8, 2019).

11 Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. *Ann Intern Med* 2004; **140**(10): 795–801.

12 CIOMS V. Current Challenges in Pharmacovigilance: Pragmatic Approaches. https://cioms.ch/wp-content/uploads/2017/01/Group5_Pharmacovigilance.pdf (accessed April 8, 2019).

13 Staffa JA, Chang J, Green L. Cerivastatin and reports of fatal rhabdomyolysis. *N Engl J Med* 2002; **346**(7): 539–40.

14 Graham DJ, Staffa JA, Shatin D, *et al*. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *JAMA* 2004; **292**(21): 2585–90.

15 Psaty BM, Furberg CD, Ray WA, Weiss NS. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: use of cerivastatin and risk of rhabdomyolysis. *JAMA* 2004; **292**(21): 2622–31.

16 Margraff F, Bertram D. Adverse drug reaction reporting by patients: an overview of fifty countries. *Drug Saf* 2014; **37**(6): 409–19.

17 Kessler DA. Introducing MEDWatch. A new approach to reporting medication and device adverse effects and product problems. *JAMA* 1993; **269**(21): 2765–8.

18 van Grootheest K, de Jong-van den Berg L. Patients' role in reporting adverse drug reactions. *Expert Opin Drug Saf* 2004; **3**(4): 363–8.

19 Inacio P, Cavaco A, Airaksinen M. The value of patient reporting to the pharmacovigilance system: a systematic review. *Br J Clin Pharmacol* 2017; **83**(2): 227–46.

20 Avery AJ, Anderson C, Bond CM, *et al*. Evaluation of patient reporting of adverse drug reactions to the UK 'Yellow Card Scheme': literature review, descriptive and qualitative analyses, and questionnaire surveys. *Health Technol Assess* 2011; **15**(20): 1–234, iii–iv.

21 Rolfes L, van Hunsel F, van der Linden L, Taxis K, van Puijenbroek E. The quality of clinical information in adverse drug reaction reports by patients and healthcare professionals: a retrospective comparative analysis. *Drug Saf* 2017; **40**(7): 607–14.

22 Watson S, Chandler RE, Taavola H, *et al*. Safety concerns reported by patients identified in a collaborative signal detection workshop using VigiBase: results and reflections from Lareb and Uppsala Monitoring Centre. *Drug Saf* 2018; **41**(2): 203–12.

23 van Hunsel F, Talsma A, van Puijenbroek E, de Jong-van den Berg L, van Grootheest K. The proportion of patient reports of suspected ADRs to signal detection in the Netherlands: case-control study. *Pharmacoepidemiol Drug Saf* 2011; **20**(3): 286–91.

24 Hammond IW, Rich DS, Gibbs TG. Effect of consumer reporting on signal detection: using

disproportionality analysis. *Expert Opin Drug Saf* 2007; **6**(6): 705–12.
25 Hazell L, Cornelius V, Hannaford P, Shakir S, Avery AJ. How do patients contribute to signal detection? A retrospective analysis of spontaneous reporting of adverse drug reactions in the UK's Yellow Card Scheme. *Drug Saf* 2013; **36**(3): 199–206.
26 World Health Organization. Safety Monitoring of Medical Products: Reporting System for the General Public. www.who.int/medicines/areas/quality_safety/safety_efficacy/EMP_ConsumerReporting_web_v2.pdf (accessed April 8, 2019).
27 Du D, Goldsmith J, Aikin KJ, Encinosa WE, Nardinelli C. Despite 2007 law requiring FDA hotline to be included in print drug ads, reporting of adverse events by consumers still low. *Health Affairs (Project Hope)* 2012; **31**(5): 1022–9.
28 Rose BJ, Fritsch BF. FDA's MedWatch program turns 20: What's new? www.pharmacist.com/fdas-medwatch-program-turns-20-whats-new (accessed April 8, 2019).
29 Muñoz MA, Delcher C, Winterstein A, *et al*. Impact of a new consumer form on adverse experience reporting to the United States Food and Drug Administration. *Pharmacoepidemiol Drug Saf* 2017; **26**(S2): 356.
30 Rogers AS, Israel E, Smith CR, *et al*. Physician knowledge, attitudes, and behavior related to reporting adverse drug events. *Arch Intern Med* 1988; **148**: 1596–600.
31 Scott HS, Rosenbaum SE, Waters WJ, *et al*. Rhode Island physicians' recognition and reporting of adverse drug events. *Rhode Island Med J* 1987; **70**: 311–16.
32 McAdams MA, Staffa JA, Dal Pan G. Estimating the extent of reporting to FDA: a case study of statin-associated rhabdomyolysis. *Pharmacoepidemiol Drug Saf* 2008; **17**(3): 229–39.
33 FDA. E2B(R3) Electronic Transmission of Individual Case Safety Reports Implementation Guide – Data Elements and Message Specification; and Appendix to the Implementation Guide – Backwards and Forwards Compatibility. www.fda.gov/drugs/guidancecomplianceregulatoryinformation/guidances/ucm274966.htm (accessed April 8, 2019).
34 US Food and Drug Administration. Guidance for Industry: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/good-pharmacovigilance-practices-and-pharmacoepidemiologic-assessment (accessed July 19, 2019).
35 McMahon AW, Pratt RG, Hammad TA, *et al*. Description of hypersensitivity adverse events following administration of heparin that was potentially contaminated with oversulfated chondroitin sulfate in early 2008. *Pharmacoepidemiol Drug Saf* 2010; **19**(9): 921–33.
36 PEW. The Social Life of Health Information, 2011. www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ (accessed April 8, 2019).
37 Edwards IR, Lindquist M. Social media and networks in pharmacovigilance: boon or bane? *Drug Saf* 2011; **34**(4): 267–71.
38 Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: a review of the opportunities and challenges. *Br J Clin Pharmacol* 2015; **80**(4): 910–20.
39 Statista. Number of social media users worldwide from 2010 to 2021 (in billions). www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (accessed April 8, 2019).
40 O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on Twitter? Mining tweets for adverse drug reactions. *AMIA Annu Symp Proc* 2014; **2014**: 924–33.
41 Patki A, Sarker A, Pimpalkhute P, *et al*. Mining adverse drug reaction signals from

social media: going beyond extraction. http://diego.asu.edu/drugstats/Publication_repository/2014/patki_biolinksig/mining%20adverse%20drug%20reaction%20signals%20from%20social%20media-going%20beyond%20extraction.pdf (accessed April 8, 2019).

42 Sarker A, Ginn R, Nikfarjam A, *et al*. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform* 2015; **54**: 202–12.

43 WEB-RADR. Recognising Adverse Drug Reactions. https://web-radr.eu/ (accessed April 8, 2019).

44 EMA. Innovative Medicines Initiative WEB-RADR Workshop Report: Mobile Technologies and Social Media as New Tools in Pharmacovigilance. www.ema.europa.eu/docs/en_GB/document_library/Report/2017/02/WC500221615.pdf (accessed April 8, 2019).

45 Brosch S. Frameworks for use of social media in pharmacovigilance. https://webradr.files.wordpress.com/2017/08/web-radr-stakeholder-event_theme-1b-ppt.pdf (accessed April 8, 2019).

46 Olsson S, Pal SN, Stergachis A, Couper M. Pharmacovigilance activities in 55 low- and middle-income countries: a questionnaire-based analysis. *Drug Saf* 2010; **33**(8): 689–703.

47 MHRA. Black Triangle Scheme. www.gov.uk/drug-safety-update/the-black-triangle-scheme-or (accessed April 8, 2019).

48 UMC. VigiBase: signalling harm and pointing to safer use. www.who-umc.org/vigibase/vigibase/vigibase-signalling-harm-and-pointing-to-safer-use/ (accessed April 8, 2019).

49 WHODrug Global. www.who-umc.org/whodrug/whodrug-portfolio/whodrug-global/ (accessed April 8, 2019).

50 MedDRA. Medical Dictionary for Regulatory Activities. www.meddra.org (accessed April 8, 2019).

51 World Health Organization. *The Importance of Pharmacovigilance. Safety Monitoring of Medicinal Products*. Geneva: WHO, 2002.

52 Bate A, Lindquist M, Edwards IR, Orre R. A data mining approach for signal detection and analysis. *Drug Saf* 2002; **25**(6): 393–7.

53 Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's Spontaneous Reports database. *Drug Saf* 2002; **25**(6): 381–92.

54 Sachs B, Riegel S, Seebeck J, *et al*. Fluoroquinolone-associated anaphylaxis in spontaneous adverse drug reaction reports in Germany: differences in reporting rates between individual fluoroquinolones and occurence after first-ever use. *Drug Saf* 2006; **29**(11): 1087–100.

55 Almenoff JS, Tonning JM, Gould AL, *et al*. Perspectives on the use of data mining in pharmacovigilance. *Drug Saf* 2005; **28**(11): 981–1007.

56 Wisniewski AFZ, Bate A, Bousquet C, *et al*. Good signal detection practices: evidence from IMI PROTECT. *Drug Saf* 2016; **39**(6): 469–90.

57 Candore G, Juhlin K, Manlik K, *et al*. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf* 2015; **38**(6): 577–87.

58 Bergvall T, Noren GN, Lindquist M. vigiGrade: a tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Saf* 2014; **37**(1): 65–77.

59 Edwards IR, Bencheikh RS. Pharmacovigilance is … Vigilance. *Drug Saf* 2016; **39**(4): 281–5.

60 Beckmann J, Hagemann U, Bahri P, *et al*. Teaching pharmacovigilance: the WHO-ISoP core elements of a comprehensive modular curriculum. *Drug Saf* 2014; **37**(10): 743–59.

61 Osimani B, Mignini F. Causal assessment of pharmaceutical treatments: why standards of

evidence should not be the same for benefits and harms? *Drug Saf* 2015; **38**(1): 1–11.
62 Jones JK. Causality assessment of suspected adverse drug reactions: a transatlantic view. *Pharmacoepidemiol Drug Saf* 1992; **1**: 251–60.
63 Benichou C, Danan G. Experts' opinion in causality assessment: results of consensus meetings. *Drug Inform J* 1991; **25**(2): 251–5.
64 Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976–1990. *Neurology* 1999; **52**: 1214.
65 Naranjo CA, Lanctôt KL. Microcomputer-assisted Bayesian differential diagnosis of severe adverse reactions to new drugs: a 4-year experience. *Drug Inform J* 1991; **25**(2): 243–50.
66 Uppsala Monitoring Centre. The Use of the WHO-UMC System for Standardised Case Causality Assessment. https://www.who.int/medicines/areas/quality_safety/safety_efficacy/WHOcausality_assessment.pdf (accessed July 19, 2019).
67 Kaufman DW, Kelly JP, Anderson T, Harmon DC, Shapiro S. Evalution of case reports of aplastic anemia among patients treated with felbamate. *Epilepsia* 1997; **38**(12): 1265–9.
68 Wysowski DK, Corken A, Gallo-Torres H, Talarico L, Rodriguez EM. Postmarketing reports of QT prolongation and ventricular arrhythmia in association with cisapride and Food and Drug Administration regulatory actions. *Am J Gastroenterol* 2001; **96**(6): 1698–703.
69 Graham DJ, Green L, Senior JR, Nourjah P. Troglitazone-induced liver failure: a case study. *Am J Med* 2003; **114**: 299–306.
70 Weber JCP. Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. In: Rainsford KD, Velo GP, eds. *Advances in Inflammation Research*. New York: Raven Press, 1984, pp. 1–7.
71 McAdams MA, Governale LA, Swartz L, Hammad TA, Dal Pan GJ. Identifying patterns of adverse event reporting for four members of the angiotensin II receptor blockers class of drugs: revisiting the Weber effect. *Pharmacoepidemiol Drug Saf* 2008; **17**: 882–9.
72 Hoffman KB, Dimbil M, Erdman CB, Tatonetti NP, Overstreet BM. The Weber effect and the United States Food and Drug Administration's Adverse Event Reporting System (FAERS): analysis of sixty-two drugs approved from 2006 to 2010. *Drug Saf* 2014; **37**(4): 283–94.
73 FDA. FDA Drug Safety Communication: FDA warns about cases of rare brain infection with MS drug Gilenya (fingolimod) in two patients with no prior exposure to immunosuppressant drugs. https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-about-cases-rare-brain-infection-ms-drug-gilenya-fingolimod (accessed July 19, 2019).
74 Berger JR, Aksamit AJ, Clifford DB, *et al*. PML diagnostic criteria: consensus statement from the AAN Neuroinfectious Disease Section. *Neurology* 2013; **80**(15): 1430–8.
75 McConeghy KW, Bress A, Qato DM, Wing C, Nutescu EA. Evaluation of dabigatran bleeding adverse reaction reports in the FDA adverse event reporting system during the first year of approval. *Pharmacotherapy* 2014; **34**(6): 561–9.
76 Southworth MR, Reichman ME, Unger EF. Dabigatran and postmarketing reports of bleeding. *N Engl J Med* 2013; **368**(14): 1272–4.
77 Graham DJ, Reichman ME, Wernecke M, *et al*. Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. *Circulation* 2015; **131**(2): 157–64.
78 Bohn J, Kortepeter C, Munoz M, Simms K, Montenegro S, Dal Pan G. Patterns in spontaneous adverse event reporting among branded and generic antiepileptic drugs. *Clin Pharmacol Ther* 2015; **97**(5): 508–17.
79 Bennett CL, Jacob S, Hymes J, Usvyat LA, Maddux FW. Anaphylaxis and hypotension

after administration of peginesatide. *New Engl J Med* 2014; **370**(21): 2055–6.

80 Weaver JL, Boyne M, Pang E, Chimalakonda K, Howard KE. Nonclinical evaluation of the potential for mast cell activation by an erythropoietin analog. *Toxicol Appl Pharmacol* 2015; **287**(3): 246–52.

81 Kotarek J, Stuart C, de Paoli SH, *et al*. Subvisible particle content, formulation, and dose of an erythropoietin peptide mimetic product are associated with severe adverse postmarketing events. *J Pharmaceut Sci* 2016; **105**(3): 1023–7.

82 Hermanson T, Bennett CL, Macdougall IC. Peginesatide for the treatment of anemia due to chronic kidney disease – an unfulfilled promise. *Expert Opin Drug Saf* 2016; **15**(10): 1421–6.

83 Mosholder AD, Gelperin K, Hammad TA, Phelan K, Johann-Liang R. Hallucinations and other psychotic symptoms associated with the use of attention-deficit/hyperactivity disorder drugs in children. *Pediatrics* 2009; **123**: 611–16.

84 FDA. MedWatch. Attention Deficit Hyperactivity Disorder (ADHD) Drug Products: Medication Guides Required to Alert Patients to Possible Cardiovascular and Psychiatric Risks. https://wayback.archive-it.org/7993/20171102213302/ https://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm107918.htm (accessed July 19, 2019).

85 World Health Organization. Reporting and Learning Systems for Medication Errors: The Role of Pharmacovigilance Centres. http://apps.who.int/medicinedocs/documents/s21625en/s21625en.pdf (accessed April 8, 2019).

86 Bencheikh RS, Benabdallah G. Medication errors: pharmacovigilance centres in detection and prevention. *Br J Clin Pharmacol* 2009; **67**(6): 687–90.

87 Newbould V, Le Meur S, Goedecke T, Kurz X. Medication errors: a characterisation of spontaneously reported cases in EudraVigilance. *Drug Saf* 2017; **40**(12): 1241–8.

88 National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). www.nccmerp.org/ (accessed April 8, 2019).

89 FDA. Drug Safety Communication: FDA approves brand name change for antidepressant drug Brintellix (vortioxetine) to avoid confusion with antiplatelet drug Brilinta (ticagrelor). www.fda.gov/Drugs/DrugSafety/ucm497942.htm.

90 Sandberg L. Follow up of UMC signals. International Society of Pharmacovigilance Annual Meeting, Pisa, 2013.

91 Thambi L, Kapcala LP, Chambers W, *et al*. Topiramate-associated secondary angle-closure glaucoma: a case series. *Arch Ophthalmol* 2002; **120**: 1108.

92 World Health Organization. 40th Annual Meeting of Representatives of National Pharmacovigilance Centres, Kampala, Uganda, 2017.

93 Russom M, Pradeep B, Zeregabr M, Fessehatzion K, Araya N. Blindness and retinal disorder associated with clomifene citrate: case series assessment *Clin Med Invest* 2017; **2**(3): 1–4.

94 Johnson W. Deployment of pharmacovigilance during mass drug administration in Sierra Leone. Paper presented at Uppsala Forum 2016, Uppsala, Sweden.

Part IIIb

Electronic Data Systems

11

Overview of Electronic Databases in Pharmacoepidemiology

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

Once hypotheses about drug-induced adverse effects are generated, usually from spontaneous reporting systems (see Chapter 10), techniques are needed to test these hypotheses. Usually between 500 and 3000 patients are exposed to the drug during Phase III testing, even if drug efficacy can be demonstrated with much smaller numbers of patients. Studies of this size would be expected to observe a single case of outcomes with an incidence of 1 per 1000 to 6 per 1000 (see Chapter 4). Given this context, postmarketing studies of drug effects must then generally include at least 10 000 exposed persons in a cohort study, or enroll diseased patients from a population of equivalent size for a case–control study. Given a study of this size, the upper 95% confidence limit for the incidence any event that is not identified would be 3 per 10 000 (see Chapter 4). However, prospective studies this large are expensive and difficult to perform. Yet such studies often need to be conducted quickly, to address acute and serious regulatory, commercial, and/or public health crises. For all of these reasons, the past decades have seen a growing use of electronic databases containing healthcare data, sometimes called "automated databases," as potential data sources for pharmacoepidemiologic studies.

Large electronic databases can often meet the need for a cost-effective and efficient means of conducting postmarketing surveillance studies. To meet the needs of pharmacoepidemiology, the ideal database would include records from inpatient and outpatient care, emergency care, mental health care, all laboratory and radiological tests (including pharmacogenomic tests that may not have been performed as part of clinical care), functional assessments, and all prescribed and over-the-counter medications, as well as alternative therapies. The population covered by the database would be large enough to permit discovery of rare events for the drug(s) in question, and the population would be stable over its lifetime. Although it is generally preferable for the population included in the database to be representative of the general population from which it is drawn, it may sometimes be advantageous to emphasize the more disadvantaged groups that may have been absent from premarketing testing. The drug(s) under investigation must of course be present in the formulary and must be prescribed in sufficient quantity to provide adequate power for analyses.

Other requirements of an ideal database are that all parts are easily linked by means of a

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

patient's unique identifier, that the records are updated on a regular basis, and that the records are verifiable and reliable. The ability to conduct medical chart review to confirm outcomes is also a necessity for most studies (unless validated algorithms for the study outcome already exist), as diagnoses entered into an electronic database may include rule-out diagnoses or interim diagnoses and recurrent/chronic, as opposed to acute, events. Information on potential confounders, such as smoking and alcohol consumption, may only be available through chart review or, more consistently, through patient interviews. With appropriate permissions and confidentiality safeguards in place, access to patients is sometimes possible and useful for assessing compliance with the medication regimen as well as for obtaining biosamples or information on other factors that may relate to drug effects. Information on drugs taken intermittently for symptom relief, over-the-counter drugs, and drugs not on the formulary must also be obtained directly from the patient.

These automated databases are the focus of this section of the book. Of course, no single database is ideal for all questions. In the current chapter, we will introduce these resources, presenting some of the general principles that apply to them all. In Chapters 12–14 of this book, we will present more detailed descriptions of those databases that have been used in a substantial amount of published research, along with the strengths and weaknesses of each.

Description

So-called automated databases have been used for pharmacoepidemiologic research in North America since 1980, and are primarily administrative in origin, generated by the request for payments, or claims, for clinical services and therapies. In contrast, electronic health record databases were developed for use by researchers in Europe, and similar databases have been developed in the US more recently.

Claims and other Administrative Databases

Claims data arise from billable interactions between patients and the healthcare system. When a patient goes to a pharmacy and gets a drug dispensed, the pharmacy bills the insurance carrier for the cost of that drug, and has to identify which medication was dispensed, the milligrams per tablet, number of tablets, etc. Analogously, if a patient goes to a hospital or to a physician for medical care, the providers of care bill the insurance carrier for the cost of the medical care, and have to justify the bill with a diagnosis. If there is a common patient identification number for both the pharmacy and the medical care claims, these elements could be linked and analyzed as a longitudinal medical record.

Since drug identity and the amount of drug dispensed affect reimbursement, and because the filing of an incorrect claim about drugs dispensed is fraud, claims are often closely audited, for example by Medicaid. Indeed, there have been numerous validity checks on the drug data in claims files that showed that the drug data are of extremely high quality, such as confirming that the patient was dispensed exactly what the claim showed was dispensed, according to the pharmacy record. In fact, claims data of this type provide some of the best data on drug exposure in pharmacoepidemiology (see Chapter 37).

The quality of disease data in these databases is somewhat less perfect. If a patient is admitted to a US hospital, the hospital charges for the care and justifies that charge by assigning diagnosis codes (until recently International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] codes) and a Diagnosis-Related Group (DRG). Hospital diagnosis codes are reasonably accurate diagnoses

that are used for clinical purposes, based primarily on the discharge diagnoses assigned by the patient's attending physician. (Of course, this does not guarantee that the physician's diagnosis is correct.) The amount paid by the insurer to the hospital is based on the DRG, so there is no financial incentive to provide incorrect diagnosis codes. In fact, most hospitals have mapped each set of diagnosis codes into the DRG code that generates the largest payment.

In contrast, however, outpatient diagnoses are assigned by the practitioners themselves, or by their office staff. Once again, reimbursement in the US does not usually depend on the actual diagnosis, but rather on the visit intensity during the outpatient medical encounter, and the resulting procedure codes indicate the intensity of the services provided. Thus, there is no incentive for the practitioner to provide incorrect diagnosis codes, but there is also no incentive for them to be particularly careful or complete about the diagnoses provided. For these reasons, the outpatient diagnoses are the weakest link in claims databases.

Some other databases are not made up of actual claims but derive from other administrative processes, such as data from US health maintenance organizations or other data sources. The characteristics of these data are similar in many ways to those of claims data, and they are discussed together as encounter-based databases in Chapter 12.

Electronic Health Record Databases

In contrast, electronic health record databases are a more recent development, arising out of the increasing use of computerization in medical care. Initially, computers were used in medicine primarily as a tool for literature searches. Then, they were used for billing. Now, however, there is increasing use of computers to record medical information at the point of care. In most instances, this is replacing the paper record as the primary medical record. As medical practices increasingly become electronic, this opens up a unique opportunity for pharmacoepidemiology, as larger and larger numbers of patients are available in such systems. The best-known and most widely used example of this approach is the UK Clinical Practice Research Datalink® (CPRD®), along with the newer database, The Health Improvement Network® (THIN®), both described in Chapter 13. As general practice databases, these contain primarily outpatient data. In addition, recently inpatient electronic health record databases are becoming available (Chapter 14).

Electronic health record databases have unique advantages. Important among them is that the validity of the diagnosis data in these databases is probably better than that in claims databases, as these data are being used to document medical care rather than just for billing purposes. When performing a pharmacoepidemiologic study using these databases, there is no purpose in validating the data against the actual medical record, since one is analyzing the data from the actual medical record. However, there are also unique issues one needs to be concerned about, especially the uncertain completeness of the data from other physicians and sites of care. Any given practitioner provides only a piece of the care a patient receives, and inpatient and outpatient care are unlikely to be recorded in a common medical record.

Strengths

Computerized databases have several important advantages, including their potential for providing a very large sample size. This is especially important in the field of pharmacoepidemiology, where achieving an adequate sample size is uniquely problematic. In addition, these databases are relatively inexpensive to use, especially given the available sample size, as they are by-products of existing administrative

systems. Studies using these data systems do not need to incur the considerable cost of data collection, other than for those subsets of the populations for whom medical records are abstracted and/or interviews are conducted. The data can be complete; for example, for claims databases, information is available on all medical care provided for covered services, regardless of who the provider was. As indicated above, this can be a problem for electronic health record databases, especially in the US, where primary care providers often do not serve as gatekeepers to specialty care. In addition, these databases can be population based, they can include outpatient drugs and diseases, and there is no opportunity for recall and interviewer bias, as they do not rely on patient recall or interviewers to obtain their data. Another advantage is that these databases can potentially be linked to other external electronic databases (e.g., death records, maternal-child records, police accident records), to expand the capabilities and scope of research. This requires the use of common identification elements (e.g., name and date of birth) and standardized semantics to allow communication across databases.

Weaknesses

The major weakness of such data systems is the uncertain validity of diagnosis data. This is especially true for claims databases, and for outpatient data. For these databases, access to medical record data for validation purposes is usually needed. This issue is less problematic for electronic health record databases; however, the validity of medication data from electronic health record databases in the United States is less certain than pharmacy dispensing data from claims databases. The addition of laboratory results data to these resources can assist in diagnosis validity, as well.

In addition, such databases can lack information on some potential confounding variables. For example, in claims databases there are no data on date of menopause, and diagnosis-based algorithms to identify smoking and alcohol abuse may have poor sensitivity, all of which can be of great importance to selected research questions. This argues that one either needs access to patients or physician records if these contain the data in question, or one needs to be selective about the research questions that one seeks to answer through these databases, avoiding questions that require data on variables which may be important potential confounders that must be controlled for.

Another major disadvantage of administrative data is the instability of the population due to job changes, employers' changes of health plans, and changes in coverage for specific employees and their family members. The opportunity for longitudinal analyses is thereby hindered by the continual enrollment and disenrollment of plan members. Another source of population instability is when patients transfer out of the system due to death or relocation. The effect of this is an inflated list with patients no longer seeking medical care. This will invalidate calculations of patient-time in studies of disease incidence, for example, because the denominator is inflated. The challenge for the investigator is to be creative in devising strategies to guard or correct for this incomplete information in the database (e.g., by performing sensitivity analysis censoring follow-up one or two years after the patient's last recorded entry in the database). Alternatively, strategies can be adopted for selecting stable populations within a particular database and, for example, by examining patterns of prescription refills for chronically used medications and restricting the study population to include only continuously enrolled patients. Of course, the largest such data system, US Medicare, suffers much less from this problem since it covers the elderly, so people never lose eligibility. Even there, however, patients can switch between fee-for-service plans and managed care plans, and the latter

may not record all healthcare which is provided (see Chapter 12).

Further, by definition, such databases only include illnesses severe enough to come to medical attention. In general, this is not a problem, since illnesses that are not serious enough to come to medical attention and yet are uncommon enough for one to seek to study them in such databases are generally of lower importance.

Some results from studies that utilize these databases may not be generalizable, for example on healthcare utilization. This is especially relevant for databases created by data from a population that is atypical in some way, such as US Medicaid data.

Finally, as noted briefly above, as an increasing number of electronic health record databases emerge in the US, to date all are problematic in that they do not include complete data on a defined population. In the US health system, unlike other countries, patients can, and often do, seek medical care from a variety of different healthcare providers at unaffiliated institutions with a nonlinked electronic health record systems. Thus, providers' electronic health records are inherently incomplete, and need to be linked to administrative data in order to be useful for quality research. This is different from the situation in, for example, the UK, where electronic health record databases are much more likely to be complete given the general practitioner gatekeeper paradigm and unique patient identifier for all healthcare services.

Particular Applications

Based on these characteristics, one can identify particular situations when these databases are uniquely useful or uniquely problematic for pharmacoepidemiologic research. These databases are useful in situations: (1) when looking for uncommon outcomes because of a large sample size; (2) when a denominator is needed to calculate incidence rates; (3) when one is studying short-term drug effects (especially when the effects require specific drug or surgical therapy that can be used as validation of the diagnosis); (4) when one is studying objective, laboratory-driven diagnoses; (5) when recall or interviewer bias could influence the association; (6) when time is limited; and (7) when the budget is limited.

Uniquely problematic situations include: (1) illnesses that do not reliably come to medical attention; (2) inpatient drug exposures that are not included in some of these databases; (3) outcomes that are poorly captured by the coding system, such as Stevens–Johnson syndrome; (4) descriptive studies, if the population studied is nonrepresentative; (5) delayed drug effects, wherein patients can lose eligibility in the interim; and (6) important confounders about which information cannot be obtained without accessing the patients, such as cigarette smoking, occupation, menarche, menopause, etc.

The Future

Given the frequent use of these data resources for pharmacoepidemiologic research in the recent past, we have already learned much about their appropriate role. As it appears that these uses will be increasing, we are likely to continue to gain more insight in the coming years, especially with the access in the US to Medicare data, and the advent in the US of the FDA's Sentinel system, exceeding 170 million individuals (see Chapter 25). However, care must be taken to ensure that all potential confounding factors of interest are available in the system or addressed in some other way, that diagnoses under study are chosen carefully, and that medical records can be obtained when needed to validate the diagnoses.

In this section of the book, Chapters 12–14, we will review the details of a number of these databases. The databases selected for review have been chosen because they have been the most widely used for published research. They are also good examples of the different types of data that are available. There are multiple others like each of them and undoubtedly many more will emerge over the ensuing years. Each has its advantages and disadvantages, but each has proven it can be useful in pharmacoepidemiologic studies.

12

Encounter Databases

Tobias Gerhard[1], Yola Moride[1], Anton Pottegård[2], and Nicole Pratt[3]

[1] *Rutgers Center for Pharmacoepidemiology and Treatment Science, Rutgers Ernest Mario School of Pharmacy, New Brunswick, NJ, USA*
[2] *Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark*
[3] *Quality Use of Medicines and Pharmacy Research Centre, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia*

Encounter databases contain electronic records of healthcare encounters for large, defined populations. They capture information on patient characteristics, prescription fills, and medical services, as part of the routine administration or reimbursement of healthcare. This is in contrast to electronic health record (EHR) databases, described in detail in Chapter 13, which are primarily intended and maintained to support patient care. Encounter data may contain records at various levels of granularity, ranging from records of individual services (fee-for-service claims) to aggregate records of care episodes (hospital discharge records). Encounter databases exist in many countries and within a number of vastly different healthcare systems. An increasing number are available for research and consequently, encounter databases have become a cornerstone of pharmacoepidemiologic research. Although they vary markedly in their specific characteristics, encounter databases share a number of defining features that warrant their discussion as a group.

While previous editions of *Pharmacoepidemiology* provided detailed information on a few select encounter databases in several dedicated chapters, this sixth edition presents this information in more general terms in a single chapter. This change in approach reflects the continued growth in the number of encounter databases used for pharmacoepidemiologic research and the often significant changes in the characteristics of individual databases over time. Rather than attempting to provide an encyclopedic description of available databases, this chapter focuses on the description of key commonalities and distinctions across encounter databases, illustrated with selected examples and supplemented by references to more comprehensive resources in the literature. The chapter now also includes a dedicated discussion of the considerations faced by researchers when evaluating the appropriateness of a specific encounter database or deciding among multiple encounter databases for their research question. The use of encounter databases for multi-database studies within distributed data networks is discussed in Chapter 25.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Description

Encounter data arise as part of the routine administration of a person's interactions with various sectors of the healthcare system. When combined, these data can be used to infer a longitudinal picture of a person's medical and treatment history. The quality of that picture, that is, its usefulness for pharmacoepidemiologic and other research, depends on the completeness and validity of the information available.

The essential attribute of all encounter databases useful for pharmacoepidemiologic research is a defined population for which healthcare services are recorded regardless of the provider or location where care is received [1]. Such databases are considered *population based* (see Chapter 17). Precise definition of the database population avoids various forms of selection bias common in nonpopulation-based studies (e.g., biased control selection in hospital-based case–control studies). Complete capture of all relevant healthcare services avoids bias from incomplete and potentially differential measurement of healthcare services (e.g., incomplete ascertainment of hospitalizations occurring in a nonparticipating healthcare system). Although representativeness of a geographic region or the general population is often desirable, it is not necessary as long as the database population is accurately defined.

While encounter databases ideally capture all healthcare services, in practice specific service types may not be captured due to the nature of the data collection process (most often due to lack of reimbursement). However, accurate qualitative description of the specific service types with lack of coverage or incomplete capture is critically important to allow evaluation of the appropriateness of the database for a given research question.

Encounter databases are maintained by a number of different entities including government agencies, insurance companies, health plans, and information services companies. The primary purpose of encounter databases is often the reimbursement of fee-for-service payment claims, and such encounter data are often referred to as claims data. In some instances, for example in US health plans with staff model delivery systems or capitated payment models, the purpose is purely administrative with no processing of payments for individual services. This distinction can be important as the accuracy and validity of data correspond to the purpose of the record. For example, claims records are routinely audited to prevent fraud and thus assure high accuracy of the data in instances where the information is directly relevant to the processing of the correct payment amount (e.g., quantity and dose of medications dispensed by a community pharmacy or type of procedure performed during an outpatient physician visit). In contrast, data elements that are not directly tied to the payment, for example the specific diagnoses associated with an outpatient visit or procedure, may be recorded with lower accuracy. In purely administrative databases, data characteristics depend on the specific data collection and quality assurance processes in place for each of the data elements.

While an ideal encounter database would capture all types of healthcare services, in practice, individual databases often lack coverage of certain service types, depending on the purpose of the database and the nature of the data collection process. The completeness of information captured in a database is a function of the types of healthcare services (data domains) included, as well as of the comprehensiveness of data capture within each domain. Encounter databases useful for pharmacoepidemiologic research typically contain the following core data domains: (1) eligibility and basic demographic information, (2) outpatient pharmacy dispensations, and (3) medical services (typically including hospitalizations; commonly also including outpatient health services).

Data domains may be maintained in separate files within a single *integrated* database (e.g., US private and governmental databases), or in multiple autonomous databases that together function as a *federated* virtual database (e.g., Nordic healthcare databases), depending on whether the data are collected and maintained by a single or multiple entities. Both integrated and federated databases require reliable linkage of an individual's records over time and between data domains. Box 12.1 summarizes commonly available data elements within the core data domains. The content of the core data domains often varies across individual databases in terms of which types of healthcare services are captured. While some databases are limited to hospital discharge data, many also capture data on outpatient office-based physician visits, outpatient clinic visits, long-term care facilities, dental, and vision. Another example of incomplete data capture within a data domain is incomplete or lack of recording of over-the-counter medication fills in prescription databases. Similar variability across databases exists in terms of access to nonencounter data, such as electronic health records, laboratory test results, diagnostic examinations, provider specialty/characteristics, vital statistics, or disease registries. Lastly, profound differences also exist in data structure and coding systems.

Because the primary purpose of encounter data is administrative, any inferences about a patient's medical history made from these data have to be carefully evaluated. Validation of encounter data, ranging from the validation of individual data elements to the validation of complex encounter data-based algorithms, is critical for rigorous pharmacoepidemiologic research with encounter databases (see Chapter 37). Validation necessitates the ability to reliably link an individual's encounter data to nonencounter data sources that serve as the external gold standard, such as electronic or paper medical records, disease registries, or survey data. Furthermore, linkage with

Box 12.1 Core data domains in encounter databases

Membership	Patient identifier, sex, age/date of birth, race/ethnicity (not universally available), zip code, dates of enrollment and disenrollment, benefits package/eligibility category (if applicable)
Medical	
Outpatient services	Patient identifier, encounter date, service location (physician office, hospital outpatient, etc.), procedure codes (e.g., CPT, HCPCS), primary and secondary diagnosis codes (e.g., ICD-10-CM), provider identifier, provider profession/specialty
Inpatient services	Patient identifier, primary diagnosis, secondary diagnoses, admission and discharge dates, length of stay, patient destination, hospital identifier. Inpatient data generally do not include information on in-hospital medication use and typically represent summaries for an entire hospital stay, resulting in some lack of detail
Pharmacy	Patient identifier, unique drug identifier (e.g., US-NDC, Nordic article number) which identifies generic name, brand name, dosage form, and strength (crosswalks may be needed for some databases while others include the individual data elements coded by the unique identifier), date dispensed, quantity dispensed, prescription duration/days supply
	Typically not recorded: indication for the prescription, inpatient drug use, over-the-counter drugs

CPT, Current Procedural Terminology; HCPCS, Healthcare Common Procedure Coding System; ICD, International Classification of Diseases; NDC, National Drug Code.

complementary nonencounter data resources or *ad hoc* data collection (see Chapter 16) is also commonly implemented in order to supplement an encounter database with variables that are required to answer a specific research question but are not available in the database, such as lifestyle factors or disease severity.

Because of their size, population-based nature, comprehensive capture of the full spectrum of healthcare encounters, and ability to rapidly assemble cohorts and identify outcomes among them, encounter databases represent a tremendous resource for pharmacoepidemiologic studies. For some research questions, encounter data may be sufficient on their own, particularly when the outcome of interest has been previously validated and data on all important confounders are available within the database. In many instances, however, validation of outcomes and supplementation with external data is necessary. In these cases, the encounter databases provide the study foundation (population base and comprehensive capture of healthcare interactions) with certain data elements critical to the study question fleshed out through linkage with additional data resources.

Attributes of Encounter Databases

Although encounter databases share a basic set of defining characteristics, they differ in numerous attributes that deserve consideration when evaluating the fit of a database to address a specific research question [2,3]. Importantly, in some databases, such as US commercial insurance databases, these attributes can be heterogeneous across individual people, as availability of supplemental data (e.g., laboratory results or ability to retrieve medical records) or even core data domains (e.g., pharmacy data) may be restricted to subsets of the full database population. In these instances, suitability of the database (e.g., in terms of sample size and representativeness) should be evaluated based on the subset of the population in a given database for which the attributes required to address the question under study (i.e., key study variables) are available rather than the database population as a whole.

Population and Coverage Period

The population captured is a critically important consideration when examining the suitability of an encounter database for the study of a specific research question. The *size of the database* is typically one of the key criteria when considering an encounter database for a specific research question, in comparison to both electronic medical record databases and alternative encounter databases. A large study population is generally necessary to ensure adequate statistical power when exposures or outcomes are rare (particularly when both are rare), effect sizes are small, and when subgroup effects or treatment effect heterogeneity are of interest. In addition, some common study designs and analytic methods may further increase the size of the database necessary to achieve adequate statistical power. For example, the new-user active comparator design results in study populations that often represent only a small fraction of the total number of users of a drug of interest during the study period [4]; restriction, a common approach to reduce confounding, can substantially decrease the size of study cohorts [5]; and instrumental variable methods are statistically inefficient compared to standard regression approaches (see Chapter 44) [6].

In addition to the size of the database, the *characteristics of the database population* have to be carefully considered. As a general rule, the population covered by an encounter database is a function of the underlying healthcare system in the respective country during the study period. Knowledge of these systems is a prerequisite for informed consideration and use of databases for pharmacoepidemiologic research. Databases in countries or regions with universal single-payer coverage, such as Taiwan, South Korea, Canadian provinces (with variations in drug benefits between provinces), and the northern European countries, generally include

the entire population and do not impose eligibility restrictions. All individuals are included and membership is maintained throughout a person's life regardless of qualifying factors such as age, employment or financial situation. As such, the characteristics of the population included in these databases are stable over time and closely track the characteristics of the population of the respective country or region as a whole. In contrast, database populations in countries or regions with less complete or more fragmented coverage, first and foremost in the US, are heterogeneous and far more complicated. The fragmentation of the US healthcare system, in particular, leads to a complex landscape for encounter databases, with different databases covering distinctly different subsets of the US population (discussed in more detail below).

Furthermore, individuals may be included in different databases at different points in time based on their personal situation (e.g., employment and state of residence), resulting in short average enrollment periods (dwell times) in any specific database environment. *Dwell time* is an important consideration particularly when the research question involves studying a long-term effect of a medicine. Similarly, when dwell time is short, it becomes increasing difficult to study new users of medicines as a lag time at the start of an individual's data capture is required to differentiate incident from prevalent medication exposure.

Lastly, the *time period covered by a database* often determines its usefulness for a given study question, depending on the start of data collection and recency of the latest available data. Studies examining trends in drug utilization over time or studies on the long-term effects of drugs, such as those with cancer as an outcome, are best served by databases with long coverage periods and a stable population. Studies of newly approved medications primarily require the most current data available. The US Medicaid Analytic Extracts (MAX) data, for example, are generally not appropriate for studies of recently approved drugs, due to an approximately three-year lag in data availability. Importantly, when studying long-term utilization trends or long-term drug effects, it is important to be aware of any changes over time in health service reimbursement and administration and appreciate their impact on drug utilization.

Services Covered and Data Completeness

For obvious reasons, medication data are a prerequisite for all encounter databases used for pharmacoepidemiologic research. Generally, these data are limited to information on medications dispensed by community pharmacies. Drugs administered during hospital stays or in long-term care units, in the emergency room, or in outpatient physician office settings are typically not included. The latter, however, can in some instances be captured through drug-specific outpatient procedure codes (e.g., drug-specific procedure codes for injection administration). In-hospital databases are discussed in Chapter 14. Over-the-counter (OTC) drug use is generally not recorded, unless OTC drugs are prescribed and specifically covered by the insurance or health system [7]. In databases for which data capture depends on a reimbursement mechanism, drug dispensings may also be missing in cases where drugs are paid for entirely out of pocket (i.e., because the cash price is lower than the required co-payment) [8], or for nonreimbursable drugs (benzodiazepines, for example, were excluded from reimbursement by Medicare Part D prior to 2013) [9].

Lastly, drug formularies, stepped therapy requirements, and prior authorization programs may impose restrictions on availability and co-payments and thus have a significant impact on use rates of individual medications and medication classes. Individual formularies may apply to an entire database population or vary widely across individuals, depending on the underlying healthcare system.

Encounter databases also vary substantially in terms of which medical services are included and, importantly, what information is captured

about these services. Most widely used encounter databases capture hospital services, including emergency departments. Hospital services are generally recorded as hospital discharge data that summarize information for an entire hospital or emergency department stay rather than provide documentation of individual services. Differences, however, exist in the granularity of these data, such as number of diagnosis fields and availability of procedure codes.

Even greater variation between databases exists in the capture of outpatient services. For example, in contrast to databases in the US, Canada, Taiwan, and South Korea, the Nordic countries do not maintain a database of outpatient office-based physician visits, though visits to outpatient hospital/specialty clinics are captured. As such, Nordic database studies of outcomes that do not result in hospitalizations or require outpatient office-based diagnoses for adjustment of confounding have to rely on medication use as a proxy for outpatient office-based diagnoses [10]. Capture of other service types, such as dental, vision, or long-term care, also depends on the database and the patient's specific insurance coverage. Lastly, particularly in the US, specific benefits such as mental health or other specialty services may be excluded ("carved out") in certain benefits packages and thus are not captured for individuals covered under these plans. For many databases, it is thus important to evaluate the availability of data on specific service types not only at the level of the database but at the individual level and over time using information on each person's benefit package.

Finally, databases differ in the information available about the patient and service provider. For example, data on the patient's race and ethnicity are generally not available in US administrative claims databases but are available in US governmental databases. Similarly, databases differ in the availability of provider specialty and identity for physician medical services as well as prescriber specialty and identity for dispensing data.

Linkage to Nonencounter Data

Many pharmacoepidemiologic research questions cannot be answered with encounter data alone. Some questions will require randomized trials (see Chapter 32) or prospective primary data collection (see Chapter 16). However, linkages to complementary sources of data may help to overcome inherent limitations of encounter data. Commonly used sources for nonencounter data include electronic or paper medical records, laboratory results, cause of death registries or autopsy records, disease or immunization registries, census data, biobanks, or survey data.

Linkage of encounter data to complementary data sources serves two distinct purposes: (1) validation of encounter-based information against an external gold standard, and (2) provision of supplementary data not available in the encounter database. Linkage to an external gold standard, ideally the medical record, for a sample of cases is particularly critical in order to facilitate outcome validation and calculation of positive predictive values (PPVs) of encounter data-based algorithms. In the absence of the medical record, validation may be performed against disease registries or patient self-report/survey. The validity of pharmacoepidemiologic drug and diagnosis data as well as approaches to the conduct of validation studies are discussed in detail in Chapter 37. The ability to retrieve medical records for outcome validation varies between databases and is often a critical factor in database selection.

Linkage to nonencounter data may also be necessary to provide supplemental information on variables that are unmeasured or poorly measured in the encounter data but necessary to adjust for confounding or appropriate restriction of the study population (e.g., indication for drug prescribing, lifestyle factors, measures of disease severity). Supplemental information such as laboratory test results or autopsy records may also be required for outcome ascertainment (e.g., Hba1c level as an outcome for a study

on the comparative effectiveness of various hypoglycemic agents).

Due to privacy restrictions that prevent the sharing or use of personal identifiers, retrieval of medical records or information obtained through direct contact with physicians or patients is generally not performed by investigators, but rather facilitated through third parties (e.g., retrieval of redacted medical records for US Medicaid and Medicare) or handled internally by employees of the participating health plans (e.g., in several US commercial insurance databases). Depending on the database, encounter and nonencounter data may be available under the same umbrella organization (e.g., linkage to EHRs in many US health plans) or require linkage to outside entities (e.g., retrieval of hospital medical records for US Medicaid beneficiaries for the purpose of outcome validation), which greatly affects the feasibility, efficiency, cost, and success rates of the linkage.

Healthcare data linkages are governed by both privacy restrictions and the availability of common linkage variables in the respective databases. Privacy regulations governing the ability to link personal health information are complex and vary between countries and database owners, and over time. When these regulations do not preclude linkage, health information databases can be linked using either deterministic or probabilistic methods [11]. Briefly, in deterministic linkage, a unique identifier or a combination of several nonunique variables available in both databases must match exactly (though the match can be implemented based on transformed versions of the variables, e.g., phonetic codes instead of names to minimize the impact of spelling errors). Deterministic linkage is most useful if reliable unique identifiers are available (e.g., US social security number) but is also achievable with combinations of multiple nonunique variables (e.g., birth dates, admission dates, and names). However, use of variables with low discriminative power and errors or missingness in the matching variable(s) will lead to a high number of overlooked (false-negative) matches.

Probabilistic linkage methods can reduce the number of overlooked (false-negative) matches by allowing imperfect matches due to partially inaccurate or missing data but in turn may produce false-positive matches. Choice of matching method thus involves a trade-off between false-negative matches (i.e., missed matches) and false-positive matches (i.e., incorrectly matched records). Simulation studies have suggested that deterministic linkage is an equally valid but less computationally intensive method for databases with low rates of missingness and error in the linkage variables [12]. However, probabilistic linkage is more accurate in error-prone data. Although often challenging, validation of linkage quality is critically important as all linkage methods are susceptible to error. The Nordic prescription database networks are examples of highly reliable linkages between encounter data and disease registries with unique identifiers [13] while the Dutch PHARMO system uses probabilistic record linkage methods [14].

Access

Access regulations, costs, and feasibility considerations vary widely between encounter databases and often have a major impact on database choice. Access may, for example, be restricted to certain researchers, such as those working in academia or governmental agencies. Some encounter databases facilitate direct access to either "off-the-shelf" or customized anonymized datasets which may be physically transferred to the researcher's institution or accessed remotely (e.g., select US commercial databases, US governmental databases, or the South Korean HIRA data), while others require in-house data analyses and thus necessitate collaborative agreements with researchers employed by the database custodian or affiliated research institutes (e.g., US health plan databases or Nordic prescription databases). Some databases are directly accessible in anonymized form but

require in-house analysis performed by the database custodian when additional "custom" linkages that require personal identifiers have to be implemented (e.g., Truven MarketScan). For studies conducted through the database custodian, it is important to not only consider the attributes of the database itself, but also the data analytic capacity and track record of the in-house research collaborators. While complexity of database structure varies between databases and studies, all work with large encounter databases requires sophisticated programming skills as well as a comprehensive understanding of database-specific details. The latter consideration can be a major advantage of collaborative arrangements that include researchers or programmers from the database custodian.

Costs of data access vary across databases and often within databases, depending on the specific characteristics of the study in question. Fees often vary by size (number of individuals) and complexity (number of files/data sources) of the requested dataset as well as by funding source (e.g., federal versus commercial funding). In-house data analysis often imposes substantial additional costs.

Application processes vary widely as well. While all databases require compliance with data privacy and security restrictions, some may also impose scientific vetting of the research plan or a justification of the benefit of the research to the public. Particularly in projects that require custom linkage with identifiable patient or provider information, close collaboration with the database custodian is needed to obtain necessary approvals and maintain confidentiality. In addition, the time required for the creation of study-specific data-cuts depends on the staffing resources and experience at the database custodian and the complexity of the required dataset. As a result, the duration from the beginning of the application process until the start of the research can vary dramatically between several weeks to multiple years.

In practice, while access considerations and familiarity with a given database are often important drivers of database choice, it is vital never to lose sight of the suitability of the database for the specific research question under study.

Selected Encounter Databases

A selection of widely used encounter databases and database types with their basic characteristics is presented in Table 12.1 and discussed below. Databases will be discussed by region and include US databases, Canadian databases, European databases, and Asian databases.

Encounter Databases in the United States

US encounter databases are arguably both the largest databases available and the most fragmented. Unlike most industrialized nations, the US does not have a uniform health system or universal healthcare coverage, resulting in databases with characteristics that differ markedly from databases in the rest of the world. In 2016, 292 million people, or 91% of the US population, had health insurance coverage, with 28 million uninsured [15]. 216 million people had coverage from private plans (68%), mostly employment-based plans (179 million; 56%). 120 million people (37%) had coverage from a governmental plan; 62 million by Medicaid (19%), 53 million by Medicare (17%), and 15 million had military coverage (5%). Note that these census data-based estimates show some inconsistencies with the reporting from the Centers for Medicare and Medicaid Services (CMS) presented later in the chapter.

Broadly speaking, most employed individuals and their dependents are covered by commercial insurance, adults 65 years and older and qualifying individuals with disabilities are covered by Medicare, and the poor and other disadvantages groups are covered by Medicaid. Furthermore, insurance coverage in the US is not mutually exclusive. In 2016, 22% of the population with

Table 12.1 Database characteristics[a]

Type	Government, US	Government, US	Health System Databases, US	Commercial Insurance, US	Government, Canada	Government, Northern Europe	Government, Asia
Examples	Medicare	Medicaid Analytic eXtract (MAX)	Kaiser, Geisinger	HealthCore, MarketScan, Optum, Pharmetrics	Saskatchewan, Quebec	Denmark, Norway, Sweden, Netherlands	South Korea, Taiwan
Networks	Sentinel	Sentinel	HCSRN, Sentinel, PCORnet, VSD	Sentinel, CNODES	CNODES	PROTECT	AsPEN
Population					Province	Country	Country
Relative size	+++	+++	++	+++	++	++	+++
Dwell time	+++	+ to ++	+ to ++	+	+++	++++	++++
Lag in availability	3–4 years	1–2 years	<1/2 year	<1/2 year	Variable	Up to 2 years	Variable
Access	Direct	Direct	In-house	In-house	<1–2 years	In-house	Variable
Retrieval of medical records for validation	Yes	Yes	Yes	Partial	No[b]	Yes	Yes for some databases
Coding, drug	NDC	NDC	NDC	NDC	AHFS	ATC	ATC
Coding, Dx	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD–8, –9 and -10	ICD-10, ICD-9
Validation	+++	+++	++++	+ to +++	++	++	++
Supplementation	+++	++	++++	+ to +++	++	+++	+++

[a] Drugs and claims only in subset.
[b] Apart from a few rare exceptions, one cannot retrieve medical charts of cases ascertained in a given study. However, can identify patients in medical records in institutions and link back to the database.

health insurance had multiple coverage types due to either switches in coverage type or to simultaneous coverage to supplement their primary insurance type.

With the exception of Medicaid programs, which generally provide prescription drug coverage for all beneficiaries, prescription drug insurance is typically provided separately from medical insurance, resulting in subgroups of patients in major databases for whom only pharmacy or medical data are available. Although pharmacy claims are recorded with high accuracy, medication dispensings can be incompletely captured in patients covered by multiple insurance programs or in instances where the co-payment is greater than the cash price of the medication [16,17]. In recent years, several large US retailers have begun to offer low-cost generic medications for as little as $4 for a monthly supply, considerably less than the average tier 1 co-payment ($11 in 2017) [8,18]. Since there is no financial incentive, pharmacies may not submit insurance claims when patients pay cash, resulting in potential underascertainment of low-cost generic medications. To date, empirical studies examining the missingness of dispensings in claims databases have reported a limited impact of such generic drug discount programs [16,17,19]. Payments rates and modalities for medical services vary widely, ranging from fee-for-service to capitated arrangements in which providers receive a fixed payment per patient per unit of time for the delivery of a specified set of services. Detailed claims data are often not available for services or patients covered by such capitated payment models as the payment amount is independent from the specific services provided.

Several large US encounter databases are available and have been widely used for pharmacoepidemiologic research [20–22]. These databases include markedly different groups of the population and often individuals with heterogeneous healthcare coverage are included within the same database. To complicate matters further, significant mobility exists between databases as changing life circumstances (loss of employment, change in employer, disability, reaching age 65/Medicare eligibility) result in changes in insurance coverage. This is often referred to as "churning" and substantially affects the average dwell time of individuals in US encounter databases [23].

US databases generally use the National Drug Code (NDC) for medication data, the Current Procedural Terminology (CPT) coding and Healthcare Common Procedure Coding System (HCPCS) for procedures, and the International Classification of Diseases, Clinical Modification (ICD-CM) system for diagnoses. The US transitioned from ICD-9-CM to ICD-10-CM on October 1, 2015 [24], which has important implications for pharmacoepidemiologic research conducted in US databases. Despite the existence of crosswalks, the performance characteristics of encounter data-based algorithms have to be demonstrated for the new coding system and studies that span the transition date will have to implement multiple coding systems in a single study. Data privacy and security of identifiable healthcare data in the US are governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [25].

US Private Insurance Databases

Most healthcare in the US is covered through private insurance, predominantly employer-based insurance. For-profit and not-for-profit insurance companies offer a wide range of plans that vary in characteristics such as premium, co-payment/co-insurance, deductibles, out-of-pocket limits, services covered, drug formularies, and provider choice. Payment systems and business models are complex and undergo continuing change over time. Because most private insurance plans are associated with the employer, many patients frequently change insurance plans due to changes in employment or when employers change their contracted insurance portfolio. Although there are hundreds of health insurance companies in the US, a relatively small number

of companies provide coverage for a majority of the privately insured population. The great majority of the privately insured population are covered by insurance systems that pay for the care provided by others [20]. Commercial insurance databases derived from these systems are some of the largest databases available for pharmacoepidemiologic research. A smaller group is covered by integrated, often not-for-profit, healthcare delivery systems that assume responsibility for preventive and therapeutic health services to a defined population, often employ group or staff model delivery systems, and frequently operate their own hospitals (e.g., Kaiser Permanente) [22]. Though typically smaller in size, the databases associated with these healthcare systems offer extensive data resources that combine encounter data with detailed clinical data resources, including EHRs and direct access to patients and providers.

Commercial insurance databases are longitudinal collections of billable healthcare interactions [20]. These databases are maintained by a variety of entities. This includes large insurance companies, often through health data analytics-focused subsidiaries (e.g., Optum Clinformatics/UnitedHealth Group [26]; HealthCore Integrated Research Database/Anthem, Comprehensive Health Insights Outcomes Data/Humana), as well as health information technology companies (e.g., Truven Health MarketScan [27], IQVIA PharmetricsPlus). Commercial insurance databases typically include several millions to tens of millions of individuals cross-sectionally and cumulatively often exceed 100 million unique patients over the life span of the database. Importantly, however, the extremely large sizes of these databases do not necessarily translate directly into the size of pharmacoepidemiologic study cohorts. Given the approximately 30% annual churn rate in commercial insurance coverage and the fact that prescription drug coverage is often separately administered or absent, only approximately 50%, 30%, and 15%, of beneficiaries with medical coverage have continuous medical and pharmacy coverage for 1, 2, and 4+ years, respectively [20].

Another important and often underappreciated feature of commercial insurance databases is the large within-database heterogeneity in data availability, completeness, quality, and ability to link member data to nonencounter data. Within a typical commercial database, members are covered by a variety of insurance products (often from multiple insurance companies), leading to substantial differences in services captured in the database. Drug formularies, which determine coverage and out-of-pocket costs for prescription drugs, for example, vary widely between plans. Similarly, a study that requires data on dental procedures would have to be limited to the subset of beneficiaries with a dental benefit during a specific time period. Completeness and quality of the claims data also depend on the payment model employed by the respective insurance products. As discussed earlier, completeness and accuracy with which services are captured may differ substantially depending on whether services are reimbursed through fee-for-service payments or capitated arrangements. Such capitated arrangements may apply to all medical coverage or be limited to specific services (e.g., specialist visits or mental health services).

The ability to validate or supplement the claims data is also often limited to subgroups of members included in the database. For example, for databases maintained by subsidies of insurance companies, data validation and supplementation may not be permitted for the (sometimes substantial) proportion of individuals in "self-funded" plans, where the employer assumes direct risk for payment and the insurance company only provides administrative services (ASO members). Similarly, the ability to identify patients and validate or supplement patient data depends on the contractual arrangements with the data sources (employers, health plans) and is generally restricted to a limited subset of the full database populations. Given the substantial

heterogeneity in multiple data attributes within and between commercial databases, thoughtful consideration of detailed information on members' individual benefit packages is critical to facilitate restriction of the study population to those for whom all necessary data elements and linkages are captured or available in the database.

Several models exist to enable research access to commercial insurance databases. Some databases are directly available in their entirety through licensing arrangements (e.g., MarketScan®), while others are solely accessible on a project-by-project basis via collaborative arrangements involving in-house programmers. Databases available for licensing are deidentified, with all personal identifiers removed, and as such do not support external linkages. Studies that require such linkages for validation or supplementation of the encounter data typically require collaboration with researchers employed by the database custodian. Such collaborations have the added advantage of tapping into the often substantial experience of the custodian research team. Most major commercial insurance providers also participate as data partners for the Sentinel System (see Chapter 25).

Integrated healthcare delivery system databases differ from commercial insurance databases in that they include a defined population whose entire spectrum of care is the responsibility of and provided by the integrated delivery system. Similar to commercial insurance databases, the delivery system databases include pharmacy dispensing data as well as encounter data on diagnoses and procedures from care delivered in both ambulatory and inpatient settings. However, because all care is provided by the delivery system, these databases also have access to full inpatient and outpatient electronic and paper medical records, and have the ability to interact with providers and patients. Although the latter features are also available for subsets of patients in many commercial insurance databases, the uniqueness of integrated delivery systems databases lies in the fact that these linkages cover the entire care received by the patient and are not limited to care received by specific practices or hospitals. Since many EHR systems include information on drugs prescribed, delivery system databases have often access to both prescription and dispensing data, which can be useful for a variety of research questions, such as questions of primary nonadherence [28]. In addition, several integrated healthcare delivery systems include affiliated research centers that maintain a variety of additional data resources such as registries for cancer, diabetes, or cardiovascular disease. Integrated health delivery systems have a long track record of pharmacoepidemiologic research, and many are consortium members in the Health Care System Research Network (HCSRN, formerly known as the HMO Research Network) and data partners for the Sentinel System (see Chapter 25) [22].

US Government

The US government funds healthcare services through several major programs, including Medicare and Medicaid, as well as the Department of Veterans Affairs Healthcare System (VA). In contrast to the VA, which is a large provider of healthcare services operating numerous hospitals, clinics, and nursing homes, Medicaid and Medicare function as payers. Both programs pay directly for services using fee-for-service arrangements, but a large and growing proportion of beneficiaries receives Medicaid (68% in 2016) [29] or Medicare (30% in 2016) [30] coverage administered by private insurance companies through capitated managed care plans. For beneficiaries covered by managed care plans, encounter data for individual services have only recently become available (Medicare) [31] or of mixed completeness and quality (Medicaid) [32] and thus research with Medicaid or Medicare data has historically been restricted to individuals with fee-for-service coverage.

The Centers for Medicare and Medicaid Services (CMS) administer Medicare and Medicaid data and facilitate access to research identifiable files for research purposes. Requests

for these data files require a research protocol and data use agreement, and are reviewed by CMS's Privacy Board. The application process is managed and supported by the Research Data Assistance Center (ResDAC) at the University of Minnesota, which provides technical assistance to researchers interested in CMS Medicare and Medicaid data. Data access requires payment of fees based on the requested population size as well as the number of data files requested, which can be provided through release of data files to investigators or remotely via the CMS Virtual Research Data Center (VRDC). A mechanism to obtain inpatient hospital and emergency department medical records corresponding to Medicare and Medicaid claims has been described and implemented [33,34]. Medicaid and Medicare data for select states or populations are also available from commercial entities (e.g., IBM Watson Health) [27].

Medicaid is a joint state/federal program intended to provide health coverage for low-income individuals. It is administered separately by each state and state-specific eligibility rules differ within federal regulations. Traditionally, the program has provided coverage limited to certain groups of low-income individuals, including pregnant women, low-income families with children, the chronically disabled, and the elderly. Following the passage of the Affordable Care Act in 2010, about one half of US states have expanded coverage to all individuals under certain income thresholds. In 2016, the average monthly enrollment in Medicaid was 70.9 million (5.7 million aged, 10.6 million blind/disabled, 28 million children, 26.7 million adults including 11.2 million adults eligible through Medicaid expansion) [30]. In 37 states, ≥50% of beneficiaries were covered through private managed care plans [29]. Medicaid coverage for eligible individuals is generally comprehensive although each state, within federally mandated parameters, administers its Medicaid program differently, resulting in variations in Medicaid coverage across the country.

Medicaid Analytic eXtract (MAX) data files include enrollment and claims data for all Medicaid enrollees in the 50 states and the District of Columbia as well as for the approximately 6.5 million (2016) enrollees in the Children's Health Insurance Program (CHIP) which serves uninsured children up to age 19 in families with incomes too high to qualify for Medicaid. MAX files have been produced since 1999 and are available per state per year [35]. MAX data are organized in five files: (1) person summary (demographic characteristics and enrollment information); (2) inpatient (inpatient hospital claims with one record per stay; procedure and diagnosis codes); (3) long-term care (e.g., nursing facility claims); (4) prescription drug (outpatient pharmacy data including national drug code, quantity dispensed, days supply); and (5) other services (e.g., laboratory and other diagnostic claims). MAX data are based on state-level data submitted through the Medicaid Statistical Information System (MSIS) and produced by CMS using extensive editing and quality control. There is a substantial lag of approximately 3–4 years between the end of a calendar year and MAX availability. Because the files are produced by state, some states may have MAX data available sooner than others. Once released, MAX data are final.

Importantly, the state reporting system is currently under transition from MSIS to Transformed-MSIS (T-MSIS). T-MSIS adds new file types (third-party liability, provider, and managed care plan data), new data elements, and modification of existing data elements [36]. One of the intentions of T-MSIS is to improve the capture and quality of encounter data for beneficiaries covered by managed care plans [37]. Data for these beneficiaries have historically been considered not to be up to research standards and have typically been excluded from most pharmacoepidemiologic research [21,32]. Given that a great majority of Medicaid enrollees are now covered under managed care plans, availability of research-quality data for this population (after

extensive quality checks and validation studies) would substantially increase the potential of MAX data as a resource for pharmacoepidemiologic research. Gaps in data capture due to periods of ineligibility are common as eligibility is typically determined monthly and changes with income and life circumstances. This issue affects individual eligibility groups differently, with more stable enrollment for those qualifying based on disability and less stable enrollment for low-income adults. Exclusion of beneficiaries without stable enrollment has been implemented based on eligibility files as well as through requirements for Medicaid encounters during specified periods before and after person-time under study.

Because Medicaid is administered at the state level, state-specific policies (e.g., opioid quantity limits or prior approval requirements) have to be considered in the research design. Medicaid and Medicare data for dually eligible beneficiaries can be linked. Such linkage is important in studies of dual enrollees since Medicaid or Medicare data alone fail to document the full spectrum of care provided to such dual enrollees [38]. Medicaid data for research are also available directly from individual states but access is often limited to researchers with established ties to the specific state Medicaid programs.

Medicare is the federal program that provides healthcare coverage for almost all people 65 years and over as well as for qualifying individuals with permanent disabilities [39]. Medicare coverage consists of four parts: Medicare Part A (Hospital Insurance), Medicare Part B (Medical Insurance), Medicare Part C (Medicare Advantage), and Medicare Part D (Medicare Prescription Drug Coverage). All parts of Medicare coverage require beneficiaries to pay deductibles and some stipulate cost sharing. Part A covers inpatient care in hospitals and skilled nursing facilities, as well as hospice. It is premium free for the great majority of beneficiaries. Part B covers physician and other outpatient services. It is an optional program that requires monthly premiums. Approximately 90% of Medicare beneficiaries enroll in Part B. Part C allows Medicare beneficiaries to enroll in private health plans that administer Part A and B benefits. The large majority of these so-called Medicare Advantage plans also include Part D benefits (i.e., prescription drug coverage). Part C plans are optional and require premiums. In 2016, 30% of Medicare beneficiaries received coverage through Medicare Advantage plans. Importantly, encounter data for Medicare Advantage beneficiaries have only recently become available through CMS (to date solely for service year 2015) [31].

Part D provides outpatient prescription drug coverage. Established in 2006, the program is administered by private companies that provide coverage through hundreds (782 in 2017) of prescription drug plans (PDPs) that differ in formulary coverage and cost sharing. Enrollment in Part D is voluntary and requires a monthly premium that varies between the individual PDPs. Medicare Part D imposes a coverage gap (doughnut hole) that requires beneficiaries to pay a substantial percentage of the cost of their medications (35% and 44% for brand name and generic drugs, respectively, in 2018) until they reach the out-of-pocket spending limit ($5000 in 2018). A large proportion of Medicare beneficiaries have some type of supplemental coverage (employer sponsored, Medicaid, so-called Medigap policies) to reduce out-of-pocket costs from cost-sharing requirements. In 2016, the average monthly Medicare enrollment was 57 million (48 million aged, 9 million disabled) [30]. 17 million beneficiaries were covered through Medicare Advantage and 41 million had a Part D benefit, including 16 million through Medicare Advantage plans [40].

Medicare data are available in several file types that are linkable to each other, as well as to Medicaid data for dually enrolled beneficiaries. File types include Master Beneficiary Summary Files (MBSFs), which include files on demographics and enrollment, chronic conditions, and cost and utilization; Institutional Claims,

which include files on inpatient services, skilled nursing facilities, and hospice; Noninstitutional Claims, which include outpatient physician claims (Carrier file) and claims for durable medical equipment; and the Part D event data file, which provides detailed prescription-level outpatient pharmacy claims. Supplementary files provide information on Part D plan characteristics, pharmacies, drugs (crosswalks from First DataBank), prescribers, and formularies.

Since prescription drug data for Medicare have become available after the establishment of Medicare Part D in 2006, Medicare, due to its large and stable population, has become one of the largest and most comprehensive resources for pharmacoepidemiologic research.

Encounter Databases in Canada

Canada, with its population of approximately 36 million, has a universal healthcare program covering all residents regardless of age or income. Program administration is the responsibility of each of its ten provinces. Physician visits, diagnostic tests, procedures (in- or outpatient), and hospitalizations are provided without payment by the patient at the point of care. Encounter data are transactional and consist of billings submitted by healthcare providers on a fee-for-service basis. A small number of physicians may have all or a portion of their activities covered by salary so the services they provide may not be included in the medical services databases. In contrast, public drug coverage programs differ among provinces; programs have been available for varying lengths of time and differ with respect to eligibility criteria as well as characteristics (i.e., copayments and deductibles). Some provinces, such as Saskatchewan and Manitoba, provide coverage for the entire population while in the others, public drug programs restrict coverage to specific segments of the population, such as the elderly, welfare recipients, or those who do not have access to private insurance plans through their employers.

Within each province, three encounter databases are available: (1) beneficiary, (2) medical services, and (3) prescription drugs. These databases are linkable through a unique patient identifier that remains unchanged over time. Additional linkage capacities are available to hospitalization databases, population health surveys [41] or province-specific disease registries [42]. Linkage of hospital charts or outpatient charts for validation of diagnoses or collection of data that are not present in the databases requires approval from the provincial information access commissioner and may not be feasible in all provinces. A number of validation studies of Canadian databases, primarily of diagnoses codes in the medical services databases, can be found in the literature but validation data remain far from comprehensive [43].

Each province maintains its own medical services encounter database, which includes all claims submitted by physicians regardless of setting (inpatient, outpatient, or emergency department) as long as the physician is paid on a fee-for-service basis. The nature of the information in the various provincial medical services databases is similar though differences exist in coding systems, such as the ICD version. For each medical service, the following information is recorded: service (date, description, location, diagnosis, and cost), provider (identifier and specialty). The vast majority of claims are submitted electronically, and the resulting medical services claims databases are populated in real time. In a few provinces, such as Nova Scotia, Manitoba, and British Columbia, mental health services, including psychotherapy, are recorded in a distinct database [44].

Unlike the medical services databases, hospitalization databases are intended for the creation of health statistics rather than for reimbursement purposes. The databases contain clinical data related to hospital discharges from acute or chronic care units, or rehabilitation centers, as well as day surgeries. With the exception of Quebec, which maintains its own hospital discharge database

(MED-ECHO), all provinces contribute to the Discharge Abstract Database (DAD) maintained by the Canadian Institute for Health Information (CIHI) [45]. The information is therefore homogeneous across provinces. In the hospitalization databases, diagnosis was coded with ICD-9-CM until 31 March 2006 and with ICD-10 thereafter. In the DAD database, information on mental health resources, cancer staging, and reproductive history was added in 2009–2010. Hospitalization databases are typically available six months after the end of the fiscal year (March 31).

Province-specific prescription drug databases record all prescription drugs dispensed in an outpatient setting to individuals covered by the public drug plan. Drugs obtained over the counter, in hospital, in long-term care units, not included in the formulary, or covered only by private insurance programs are not usually included in the database. One exception is PharmaNet in British Columbia that links all pharmacies to a central data system. Every prescription dispensed in the outpatient setting is recorded regardless of coverage; hence, it includes medications covered by the public drug plan and private insurance programs, as well as those acquired out of pocket. Drugs are coded according to the Canadian-specific Drug Information Number (DIN) as well as the American Hospital Formulary Service (AHFS). For each dispensing, the following information is recorded: drug (date of dispensing, drug name, dose per unit, mode of administration, prescribed duration [not recorded in Saskatchewan], cost including dispensing fees), pharmacist (identifier, pharmacy location), and prescriber (identifier, specialty). Indication for a drug prescription is not recorded in any of the dispensing databases. While data and coding systems are similar across provinces, inclusion of individual drugs in the formulary and type of listing (general or restricted) may vary. For each patient, the years of entry and exit from the drug program are available in the beneficiary database. This is important information for studies that include segments of the population whose membership in the drug program may be transitory, such as membership based on income or access to private insurance programs.

Only seven of the 10 Canadian provinces make prescription data available for pharmacoepidemiologic research. Approximately half of these databases are accessible through custodians located in a university setting while the other half are accessible through provincial government agencies. In addition to the drug databases, custodians also act as a repository for other provincial databases and are responsible for their linkage.

Database access varies across provinces. Some provinces (Saskatchewan, Quebec, Nova Scotia) provide raw anonymized datasets to researchers (from academic or industry settings) while others (Ontario, BC) require data to be analyzed in-house by specific research organizations. To maintain confidentiality of the data, no patient, healthcare provider (including pharmacist), or institution identifiers are transmitted to researchers. Additional restrictions are in place in individual provinces. For example, in Quebec only a random sample of approximately 75% of the population eligible for a given study (capped at a maximum of 125 000 eligible patients) may be obtained, and no birthdates are transmitted. Exceptions can be granted through a request to the Provincial Access to Information Commission, which substantially increases the delay in data extraction.

Although Canadian encounter databases are much smaller than US encounter databases, their greatest advantage is that they include a stable population, thereby allowing longer follow-up periods. This is, for example, illustrated through a study on benzodiazepines and Alzheimer's disease, in which a 10-year follow-up was available [46]. The time required for database extraction varies across provinces, ranging from 10–20 weeks to one year, more if a request to the Provincial Access to Information Commission is required.

Encounter Databases in Europe

Nordic Prescription Databases

The Nordic countries (Denmark, Iceland, Norway, Sweden, and Finland) have tax-supported universal health coverage. All citizens (a combined population of over 25 million people ranging from ~300 000 in Iceland to more than 9 million in Sweden) are provided with unrestricted access to health services including partial or complete reimbursement of medications.

Pharmacies electronically submit information on dispensed prescriptions to national databases without a requirement for informed consent by the patient (available since 1994 in Finland and Denmark, 2004 in Norway, 2005 in Sweden, and 2006 in Iceland) [47]. Unique civil registration codes facilitate unambiguous linkage to various national databases using a central patient router file. Linkable national databases include but are not limited to hospital discharge databases, laboratory data including results, pathology databases, medical birth databases, cancer registries, and cause of death databases, as well as census data, health surveys, biobanks, and patient records. Together, these databases create a federated database network that provides exposure information from the prescription database as well as patient and clinical outcome data from the patient router file and multiple linked autonomous databases.

The prescription databases largely include similar data elements with slight variations between countries. Besides a patient identifier (which also encodes birth year and sex), data include drug data (dispensing date, Nordic article number, a unique identifier similar to the NDC code used in the US, ATC classification, quantity dispensed in defined daily doses), a prescriber identifier (which can be linked to prescriber data such as basic demographics, profession, specialty, practice site), and pharmacy data (name and location). OTC drugs are not included unless they are obtained via prescription. Importantly, some drugs that are also available OTC are used primarily via prescription, to ensure reimbursement [7]. Besides the difference in the age of the databases, the most noticeable difference is the fact that nonreimbursed drugs are not covered by the Finnish database.

Outcome data are primarily based on national hospital discharge databases (registries). While comparable, some differences exist in the age of the patient databases, with the Finnish database dating back to 1969 [48], followed by the Danish (1977) [49], Swedish (1987) [50], and Norwegian registry (2008) [51]. Numerous other databases including cancer, birth, and death, together with pathology and laboratory results, further complement the dataset. Importantly, no large-scale data are available that provide details regarding general practice visits or other nonhospital health services. This is often referred to as a lack of "outpatient" data. However, this term can lead to misunderstandings in the context of the Nordic healthcare model. All hospital databases cover activities within hospital outpatient clinics, and as such all specialized care is covered. However, in all Nordic countries, general practice physicians serve as gatekeepers to specialized care (including both hospital and private practicing specialists). Detailed data, such as diagnoses or laboratory data, are not available. However, data on contacts (without specification for the reason for such contacts) can be obtained.

Rules governing data access vary between the Nordic countries, but generally require collaboration with local researchers. Access to Danish prescription data is particularly restrictive. Consequently, data from the Danish National Prescription Registry [52] cannot leave the data havens provided by Danish authorities. For multinational studies involving Danish individual-level prescription data, pooled analyses require data to be transferred to, for example, Statistics Denmark [53] or metaanalysis techniques to be applied to obtain pooled estimates [54]. Other sources of Danish prescription data are not restricted in the same way, but either only offer local coverage [55,56] or only provide

data on reimbursed prescriptions and only cover more recent years [57].

Other European Encounter Databases

Pharmacy-based federated database networks also exist in The Netherlands (PHARMO) [14] and Scotland (Tayside MEMO) [58]. These networks are limited to specific regions of their respective countries and have the ability to link to a number of databases that provide outcome and confounder information similar to those in the Nordic countries. In addition, integrated encounter databases are available in France [59] and some regions of Italy (Lombardy, Tuscany) [60]. The French national claims database, SNIIRAM, captures data for more than 66 million individuals (~98% of the French population) regardless of socioeconomic or employment status. It captures encounter data on outpatient visits, dispensed medication, procedures, chronic conditions, hospital admission diagnoses and procedures, and date of death. Data access, however, is complex.

Encounter Databases in Asia

There are many encounter databases available across the Asia-Pacific region. Many of these are population-wide databases due to the prominence of nationwide healthcare coverage in these countries. For example, South Korea and Taiwan both have single-payer, universal government-run health insurance systems that predominantly operate on a fee-for-service basis and have established national research databases. The National Health Insurance Databases of South Korea and Taiwan are the most well-established and widely used Asian encounter databases. Similar to encounter databases in the US, Canada, and Europe, they capture patient demographic information, medical (in- and outpatient) services and prescription and dispensing data. Encounter databases also exist in Australia and Japan [61]. In Australia, the commonwealth government maintains a dataset of dispensing of subsidized medicines under the Pharmaceutical Benefits Scheme (PBS) and medical services under the Medicare Benefits Schedule (MBS) [62]. A 10% sample of these data, linked longitudinally, is available and has been used for research [63,64]. Additionally, an encounter database of services provided to Australian veterans is maintained by the Australian Department of Veterans Affairs (DVA). These data include all prescriptions dispensed, medical services claimed and hospital visits attended by the veterans, their dependents, and spouses. The DVA data have been used widely for research [65,66].

One of the advantages of databases across the Asia-Pacific region is the consistency of coding systems. For example, encounter databases in South Korea, Taiwan, and Australia all use ATC codes to identify individual medicines and all but Taiwan use ICD-10 codes to identify diagnoses. This allows for comparisons of similar products across different countries without the need to map individual country-specific codes. This has allowed cross-national studies to be conducted using a distributed network approach through the Asian Pharmacoepidemiology Network (AsPEN) [67]. Pharmacoepidemiologic studies using Asian databases have historically been limited due to restrictions in the accessibility of these data. One study found that of 54 encounter databases across the Asia-Pacific region, very few allowed access to raw data [68]. Databases in Australia, Taiwan, and Japan, for example, were considered as having a high level of data accessibility, while South Korea had a medium level and Thailand, China, Malaysia, and Singapore had a low level of accessibility. The level of accessibility can differ for individual databases within the same country; some databases may require a local researcher to access data while others do not provide raw data with only summary-level data available for researchers.

Taiwanese National Health Insurance Research Database

Established in 1995, the National Health Insurance (NHI) program of Taiwan covers approximately

23 million individuals, more than 99% of the country's population [69]. The NHI maintains the National Health Insurance Research Database (NHIRD), which is accessible for research. The NHIRD includes but is not limited to patient demographics, prescription and dispensing data, outpatient visits, hospitalizations, and dental care. Data are updated biannually. The NHIRD can be linked to a number of external national databases through a unique and universal personal identification number. Databases available for linkage include numerous registries (birth, death, immunization, cancer, reportable infectious diseases, suicide), population-based screening programs (various cancers, myopia, urine, newborns) as well as regular examinations in school children. Strict procedures for data access and human subject review are in place to assure protection of confidentiality and data security.

South Korean Health Insurance Review and Assessment Data

South Korea has provided universal health coverage since 1989. In 2000, all health insurance systems were integrated into a single national system, creating the National Health Insurance Service (NHIS) and the Health Insurance Review and Assessment Service (HIRA). All healthcare providers are covered under the NHIS and are, with a few exceptions, reimbursed on a fee-for-service basis. Claims are electronically submitted by providers to the HIRA for reimbursement and form the basis for the HIRA database, which contains healthcare utilization and prescribed medications for approximately 50 million individuals [70]. Use of the database was initially limited until it became publicly available for research in 2009.

The HIRA research data include beneficiary ID, basic demographics, procedures, diagnostic tests, all diagnosis received by the beneficiary (coded in KCD6, the Korean Standard Classification of Disease Version 6, which is closely based on the ICD-10 system), in- and outpatient prescriptions (including brand name, generic name, prescription and dispensing date, duration, dose, and route of administration), as well as provider ID and characteristics. Validity of diagnosis data in the HIRA database has been shown to vary according to the severity of the condition (with greater validity for more severe conditions) and the care setting (with higher validity for inpatient than outpatient diagnoses) [71]. HIRA data are available to researchers in academia and government agencies and for those in the private sector such as pharmaceutical companies and medical device companies but access requires in-person consultation at the HIRA and submission of a study proposal. Once approval is given, tailored data extracts with encrypted ID information for protection of privacy are uploaded in a remote access system accessible only by the individual researcher for the study. Importantly, HIRA data are currently available only for a five-year period beginning from the current year although plans exist to expand this period to 10 years.

Strengths

Encounter databases have a number of strengths in comparison to other data sources for pharmacoepidemiologic research, which explain their broad representation in the literature.

First, automated healthcare databases facilitate the rapid and cost-efficient assembly of extremely large cohorts of patients and provide data on drug exposures, health outcomes, and potential confounding factors. Encounter databases, in particular, are the largest available population-based healthcare databases. Several of the databases discussed in this chapter cumulatively include more than 100 million individuals and provide the ability to rapidly assemble cohorts that are substantially larger than analogous cohorts from EHR databases or *ad hoc* data collection.

Encounter databases thus are uniquely able to address research questions that require the largest possible study sizes. The following example

illustrates the differences in cohort sizes for the same study in selected encounter and EHR databases. Filion and colleagues examined proton pump inhibitors and the risk of hospitalization for community-acquired pneumonia among new users of NSAIDs, aged ≥40 years in multiple databases within the Canadian Network for Observational Drug Effect Studies (CNODES) [72]. The respective sizes of study cohorts assembled using a common protocol and allowing multiple cohort entry dates for a single patient were approximately 2.2 million for MarketScan, 1.5 million for the combined Canadian provincial databases, and 0.6 million for the UK GPRD, the largest population-based EHR database. The MarketScan cohort was more than 3.5 times larger than the GRPD cohort, despite not including data on ≥65 year olds who made up around 35% of the total study population.

Second, because encounter databases are population based and provide a comprehensive capture of covered healthcare encounters regardless of the provider, they can support the full range of epidemiologic study designs including cohort, nested case–control, and self-controlled designs. While this strength is shared by a number of other population-based automated databases, it is a critical limitation to nonpopulation-based data sources such as EHR databases of individual institutions or health systems.

Third, many encounter databases facilitate systematic or *ad hoc* linkage to nonencounter data resources, including electronic or paper medical records, disease registries, laboratory results, or patient and provider surveys. Such linkages can support validation of study outcomes and allow supplementation of encounter data with variables such as laboratory results or lifestyle data. In ideal circumstances, such linkages thus provide the ability to take advantage of the size and population-based nature of encounter data, while also accruing the advantages of higher data quality and greater clinical detail available from data sources such as EHRs, disease registries, or patient and provider surveys. Importantly, however, linkage ability and quality vary substantially between individual encounter databases and have to be carefully considered for each study question.

Fourth, many large encounter databases are broadly representative of nations, regions, or particular health systems. As such, they can often serve an important role in facilitating health services and health policy research. Many include very stable populations that facilitate assessment of long-term safety effects and long-term trends in treatment practice and quality. Further, encounter databases from countries or regions with universal health coverage – by definition – are free from selection bias as inclusion in the database is universal.

Fifth, for encounter data generated from fee-for-service payment claims, data elements that directly pertain to the payment amount are subject to auditing and considered highly accurate. This is true for procedure claims (type of procedure performed) [73] as well as for pharmacy claims (date, drug, and quantity dispensed) [74]. Importantly however, the accuracy of procedure data primarily relates to the occurrence of the procedure billed while the accuracy of the clinical indication associated with the procedure may be substantially lower. For example, a validation study that used specific surgical procedure codes in Medicaid data as part of an algorithm to identify cases of hip fracture found in medical record review that while all of the procedures billed for were actually performed, some of the procedures were used to correct orthopedic conditions other than hip fracture [75]. A further advantage of pharmacy data compared to prescription data recorded in EHR databases (see Chapter 13) is the fact that prescription dispensings are one step closer to ingestion than what was prescribed and thus are subject to a lesser degree of exposure misclassification [76]. The accuracy of encounter data generated by administrative processes not

related to payment is less well established and likely to vary depending on the existence and rigor of quality assurance processes.

Sixth and last, data capture processes in encounter data are automated and independent of the study question and hypothesis, greatly diminishing the likelihood of recall or assessment biases.

Limitations

Encounter databases are primarily intended and maintained for payment or other administrative purposes, and therefore are subject to important limitations when used for research.

First, one of the greatest concerns when using encounter databases for pharmacoepidemiologic research is the uncertain validity of diagnostic information (see Chapters 11 and 37) [49]. While these concerns apply to all diagnostic encounter data, they are amplified for diagnoses recorded in the outpatient setting where diagnosis is typically not directly linked to a particular level of payment. It is thus critically important for all encounter-based research to validate diagnostic data (for both outcomes and important confounders) against external gold standards such as the medical record or disease registries. These gold standards, of course, may not be correct either when compared to research-grade diagnoses as employed by randomized controlled trials.

Second, encounter data lack clinical detail such as markers of disease severity (e.g., blood pressure, ejection fraction) and lifestyle factors (tobacco and alcohol use, body mass index, physical activity). Oftentimes, data elements are available (e.g., diagnostic codes for obesity or smoking status) but of extremely low sensitivity. For example, a study using data from the National Health and Nutrition Examination Survey to validate diagnosis of obesity in Medicare claims found that claims-based diagnostics codes fail to identify a great majority of patients with obesity (sensitivity of 18%) [77]. Though still far from perfect, clinical details such as disease severity and lifestyle factors are generally better captured by paper or electronic medical records. Because such clinical detail is often critical for confounding adjustment, methods that minimize unmeasured or residual confounding (self-controlled designs, active comparator new-user designs, instrumental variable analyses, propensity score calibration) are of great importance to encounter-based pharmacoepidemiologic research (see Chapter 43).

Third, while limitations of encounter databases can often be overcome by facilitating linkage to nonencounter data such as EHRs, disease registries, or laboratory results, such linkages are typically time-consuming and costly and, in many cases, only available to subsets of the database population. Further, when compared to population-based EHR databases, the resulting linked/enriched encounter data typically remain less comprehensive, and validation is often restricted to small samples often with poor response/retrieval rates.

Fourth, in certain situations, medication dispensing information may not capture data for specific drugs or drug classes. This may include drugs excluded from reimbursement, drugs that are primarily obtained over the counter, as well as low-cost generic drugs that are paid for out of pocket because the cash price is lower than the required co-payment. This may result in misclassification of exposure, such that some patients will appear not to be exposed to a medicine when in fact they were. Nonreimbursable drugs as well as low-cost generics are often better captured in EHR databases, which contain information on all prescriptions written. However, the disadvantage of prescription information is that not all prescriptions will be dispensed and will result in misclassification of exposure, such that some patients will appear to be exposed to a medicine when in fact they were not.

Fifth and last, due to the fragmentation of the US healthcare system, many large US encounter databases lack representativeness of the general population and feature significant turnover and short dwell times (e.g., US private insurance databases, MAX) [20,21,78].

Particular Applications

Encounter databases have been used in thousands of pharmacoepidemiologic publications, many of which have shaped clinical medicine or regulatory decision making. These databases have supported work across a wide spectrum of areas including drug safety, comparative effectiveness, drug utilization and health services research, methods and validation, as well as pharmacoeconomics. Descriptions of numerous specific applications of individual databases can be found in the 5th edition of *Pharmacoepidemiology* [14,20–22,45].This section outlines some typical activities involved in encounter database studies and presents some of the considerations in choosing the optimal encounter database when multiple options are available or assessing the suitability of a specific database for a given research question.

Typical Activities Involved in Studies Using Encounter Databases

Although encounter databases vary in data structure, coding schemes, and numerous other specifics, a number of activities are typical across all such databases [20]. Virtually all pharmacoepidemiologic studies of encounter databases require *linkage of records between data files and over time*. Records from different data domains, such as membership, outpatient services, inpatient services, and pharmacy, are linked so that an individual's entire set of encounters over the study period can be available for analysis. Another ubiquitous step in the conduct of pharmacoepidemiologic studies involves the *aggregation of drug, diagnosis, and procedure codes into meaningful study variables*. Exposures, outcomes, potential confounders, and inclusion/exclusion criteria for study are defined via code lists using drug, diagnosis, and procedure codes, or combinations thereof. These code lists are typically study and database specific using the coding schemes utilized by the respective database and drugs approved and available for the study population during the study period. It is often desirable to use previously validated algorithms for the definition of study outcomes and important confounding variables. Such algorithms often combine diagnostic codes, drug codes, and procedure codes for more accurate measures of disease (see Chapter 37).

Together with demographic information, these study-specific variables (e.g., drug classes, disease states) facilitate the *creation of the study population*. Study populations often consist of (new) users of specific drugs or drug classes within individuals who meet specific inclusion and exclusion criteria based on their encounter-derived medical history. Once the study population is identified in the dataset, analytic plans often specify the construction of longitudinal histories. Exposure, occurrence of outcome events, and presence of confounding factors are measured over time, typically in temporal relation to the study's index date. This facilitates the assessment of exposure periods and person-time at risk, and allows calculation of incidence rates and measures of association. If additional data not available in the encounter database are required, complementary information may be gathered through linkage to electronic medical records, data obtained directly from patients or their physicians from surveys, retrieval of paper medical records, or data routinely collected in disease, immunization, or national vital registries.

Deciding Between Individual Encounter Databases

Database choice or evaluation of suitability of a single database should involve consideration of all database attributes relevant to the research question under study [3]. Some of the key attributes that differentiate individual encounter databases are shown in Table 12.1 and discussed below.

Target Population

The database should capture a large and representative sample of the target population (e.g., patients exposed to a particular drug) to adequately address the study question. For example, Stroup and colleagues aimed to examine the effectiveness of initiating treatment with either clozapine or a standard antipsychotic among adults with evidence of treatment-resistant schizophrenia using national US Medicaid data [79]. On first glance, this might not be an obvious choice as the adult Medicaid population is highly selective and often transient. However, Medicaid covers approximately two-thirds of all US adults with schizophrenia because most patients with severe schizophrenia qualify for disability [80]. In addition, because these individuals qualify for Medicaid because of disability rather than because of their economic condition, they are typically stably enrolled without breaks in coverage. While non-US encounter databases might have provided similarly large numbers of stably enrolled patients with schizophrenia, the authors sought a US database because of the pronounced differences in psychiatric treatment practice between US and most other countries. The study was conducted as a 1:1 propensity score matched cohort study and found that clozapine-treated patients compared to patients treated with a standard antipsychotic had a decreased risk of psychiatric hospital admission (hazard ratio 0.78, 95% confidence interval (CI) 0.69–0.88) but an increased risk of diabetes mellitus (hazard ratio 1.63, 95% CI 0.98–2.70).

Database Size

The database should be large enough to provide sufficient power to answer the research question, that is, to detect a meaningful difference between treatment groups (should a difference truly exist). This assessment should be based not on the size of the overall database but rather the size of the actual study cohort, that is, the cohort after exclusion of individuals for whom required data elements are unavailable (e.g., after exclusion of individuals under capitated payment plans), and after application of inclusion and exclusion criteria (e.g., sufficient uninterrupted baseline period).

A study by Shin et al. aimed to determine the risk of cardiovascular conditions in children and adolescents with ADHD associated with use of methylphenidate [81]. As the outcome was rare, the South Korean HIRA database of over 50 million participants was used. From this large population database, 144 258 patients aged less than 18 with a diagnosis of ADHD were retrieved. Of these, 114 657 were new users of methylphenidate and 1224 had an incident cardiovascular event. Due to the rare outcome, a self-controlled case series design was used which, compared to other designs, has the advantage of requiring fewer patients for similar power (see Chapter 43).

Ability to Validate Outcomes

Because encounter data are primarily collected for administrative purposes, the ability to validate or adjudicate outcome definitions derived from these data is essential for pharmacoepidemiologic studies. Outcome validation should generally be performed as part of any encounter-based study unless the outcome measures have previously been validated for the database. However, the ability to validate outcomes, through reliable linkage to external gold standards such as the medical record or disease registries, varies markedly between databases and is often a major consideration for database selection.

Lo Re and colleagues, for example, conducted a series of postauthorization safety studies to examine the safety (hospitalization for major adverse cardiovascular events, acute kidney injury, acute liver failure, infections, and severe hypersensitivity events) of saxagliptin compared to other oral antidiabetic drugs in patients with type 2 diabetes [82,83]. The studies were conducted separately in two EHR databases (Clinical Practice Research Datalink, The Health Improvement Network) and two encounter databases (Medicare, HealthCore Integrated Research Database). One of the requirements for the choice of encounter databases in this study was the ability to obtain inpatient medical records for outcome adjudication. Using a new-user active comparator cohort design, the study found no evidence of increased risk of any of the outcome events within any of the four databases.

Other outcomes are notoriously undercoded in encounter data and require development of custom algorithms. For example, using data from Quebec, Moride et al. developed and validated a case detection algorithm for suicide attempts in youth through a review of medical charts [84]. The following algorithm was used: diagnostic code of injury or intoxication with a location of service in the ED, followed by a psychiatric consult or a psychiatric diagnosis (psychiatric diagnoses consisting of depression, eating disorder, schizophrenia, ADHD, substance abuse, others) within two days of the ED visit. This algorithm had a sensitivity of 70% and a specificity of 97.6%.

Availability of Nonstandard Encounter Data

While all encounter databases provide information on medical services and prescription drugs, studies often require encounter data on services that are not universally available in all databases. For example, Gupta and colleagues examined opioid prescribing practices among US dentists from 2010 to 2015 using the MarketScan database [85]. Because dental services are not captured for all individuals in the database, the study population was appropriately restricted to those with simultaneous enrollment in a medical and a dental plan.

Ability to supplement with non-encounter data: Studies using encounter data may require clinical detail not available from encounter data often for the purpose of confounding adjustment or to supplement outcome identification. The ability to perform linkages that allow enrichment of the dataset with non-encounter data is thus vital and often a decisive consideration in choosing a study database. For example, Huybrechts and colleagues examined the comparative mortality risk of individual antipsychotics in elderly nursing home residents using data for US nursing home residents dually eligible for Medicaid and Medicare [86]. Clinical variables such as cognitive function or behavioral symptoms of dementia are important potential confounders but poorly measured in encounter databases. Linkage to the Minimum Data Set (MDS, available from CMS), a federally mandated health assessment tool used in nursing homes that captures information on physical, psychological, and psychosocial functioning, active clinical diagnoses, health conditions, treatments, and services, allowed the inclusion of these important covariates into the study. Using a propensity score-adjusted new-user cohort design, the authors showed that compared to initiators of risperidone, initiators of haloperidol had an increased mortality risk and initiators of quetiapine had a decreased mortality risk.

As another example, a Swedish-Danish study investigated the risks associated with being admitted to an emergency department with suspected poisoning, most often psychotropics or analgesics. Leveraging the ability to link data on admissions and prescription fills to a dataset including detailed ECGs on those admitted to the hospital, they could estimate not only the occurrence of QTc prolongation within the population but also to what extent QTc prolongation as a marker was associated with 30-day mortality [87].

The Future

Pharmacoepidemiologic research with encounter databases has become more and more widely used and involves an increasing number of databases in a growing number of regions of the world. This trend is expected to continue, particularly as encounter databases become available in regions for which currently no data are available. In addition, the following three major factors are likely to shape the future of encounter databases: (1) advances in information technology (IT), (2) privacy regulations, and (3) changing healthcare systems. Advances in IT will continue to expand the boundaries of data storage and processing, and increasingly facilitate linkages with new and more complex sources of data, including biomarkers, social media, web searches, and around-the-clock biometric information from wearables. In addition, automated tools for data visualization and analysis of health data are becoming more accessible.

The potential for rapid development of progressively complex, detailed, and complete data resources is likely to be counteracted by increasingly strict regulations governing data privacy. These regulations will vary substantially between countries and are likely subject to rapid change.

Last, and maybe most importantly, encounter-based data are a secondary byproduct of administrative systems, created to support the local healthcare system; research applications are secondary uses. As such, encounter-based healthcare data will continue to be subject to changes in the healthcare systems that generate the data. Again, these changes are likely to vary drastically between countries and over time.

For example, the US healthcare environment is undergoing enormous transformation. Historically, healthcare providers in the US have been paid using a fee-for-service approach, where providers bill health insurance companies for the cost of the services they provide, generally justifying those bills with diagnoses. These paid claims represent the core of these encounter databases. However, the net result of this approach is that the more providers do, the more they are paid, which may result in overservicing and wasted resources. The result has been a large incentive to increase utilization, and rapidly increasing costs in the US for providing healthcare, made worse by an aging population. Under this model, the levels of expenditure are unsustainable. This has led to a shift from a fee-for-service model to a "per patient per month" payment system, so-called "population health", which of course switches the incentive to providing less care. In order to attempt to address that, incentives are being put in place to ensure that people are not receiving *too little* care, referred to as "value health". The US is in the middle of this transition now, varying greatly in different parts of the country. However, in response, there has been a remarkable consolidation of physician practices, hospitals, etc., in order to achieve sufficient scale to create the needed extensive and costly data infrastructure, and to assume the large risk associated with population health. Many other initiatives are under way as well, to limit the increasing costs of medical care. The results will likely be large changes over the next few years in the data as part of US encounter databases.

Encounter-based data are an important resource for pharmacoepidemiologic research. These data are comprehensive and often have a high level of quality as they are collected for payment purposes. As these data are generated for purposes other than research, consideration of their applicability, completeness, and generalizability needs to be carefully weighed against their convenience. As with any data source, careful consideration should be given to the issues of bias and confounding (see Chapter 3) which are not problems diminished by the increased size of the database.

References

1 Szklo M. Population-based cohort studies. *Epidemiol Rev* 1998; **20**(1): 81–90.

2 Strom BL. How should one perform pharmacoepidemiologic studies? Choosing among the available alternatives. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 364–76.

3 Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, LoCasale R. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012; **21**(1): 1–10.

4 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**(9): 915–20.

5 Schneeweiss S, Patrick AR, Sturmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–42.

6 Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 537–54.

7 Schmidt M, Hallas J, Friis S. Potential of prescription registries to capture individual-level use of aspirin and other nonsteroidal anti-inflammatory drugs in Denmark: trends in utilization 1999–2012. *Clin Epidemiol* 2014; **6**: 155–68.

8 Choudhry NK, Shrank WH. Four-dollar generics – increased accessibility, impaired quality assurance. *N Engl J Med* 2010; **363**(20): 1885–7.

9 Centers for Medicare & Medicaid Servicers. Transition to Part D Coverage of Benzodiazepines and Barbiturates Beginning in 2013. www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/BenzoandBarbituratesin2013.pdf (accessed April 8, 2019).

10 Mikkelsen KH, Knop FK, Frost M, Hallas J, Pottegard A. Use of antibiotics and risk of type 2 diabetes: a population-based case-control study. *J Clin Endocrinol Metab* 2015; **100**(10): 3633–40.

11 Mason CA, Tu S. Data linkage using probabilistic decision rules: a primer. *Birth Defects Res A Clin Mol Teratol* 2008; **82**(11): 812–21.

12 Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J Biomed Inform* 2015; **56**: 80–6.

13 Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014; **29**(8): 541–9.

14 Herings RMC, Pedersen L. Pharmacy-based medical record linkage systems. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 271–86.

15 Barnett JC, Berchick ER. Current Population Reports, P60-260, Health Insurance Coverage in the United States: 2016. Washington, DC: US Government Printing Office, 2017. www.census.gov/content/dam/Census/library/publications/2017/demo/p60-260.pdf (accessed April 8, 2019).

16 Cepeda MS, Fife D, Denarie M, Bradford D, Roy S, Yuan Y. Quantification of missing prescriptions in commercial claims databases: results of a cohort study. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 386–92.

17 Zhou L, Stearns SC, Thudium EM, Alburikan KA, Rodgers JE. Assessing Medicare Part D claim completeness using medication self-reports: the role of veteran status and Generic Drug Discount Programs. *Med Care* 2015; **53**(5): 463–70.

18 Claxton G, Rae M, Long M, Damico A. Kaiser Family Foundation, Employer Health Benefits, 2018 Annual Survey. http://files.kff.org/

attachment/Report-Employer-Health-Benefits-Annual-Survey-2018 (accessed April 9, 2019).
19 Roberto PN, Stuart B. Out-of-plan medication in Medicare Part D. *Am J Manag Care* 2014; **20**(9): 743–8.
20 Seeger J, Daniel GW. Commercial insurance databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 189–208.
21 Hennessy S, Palumbo Freeman C, Cunningham F. US Government claims databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 209–23.
22 Andrade SE, Raebel MA, Boudreau D, *et al*. Health maintenance organizations/health plans. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 163–88.
23 Austic EA, Lawton E, Riba M, Udow-Phillips, M. Insurance churning. Cover Michigan Survey 2015. Ann Arbor, MI: Center for Healthcare Research and Transformation, 2016.
24 International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). Centers for Disease Control and Prevention; 2016. www.cdc.gov/nchs/icd/icd10cm.htm (accessed April 9, 2019).
25 Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191 (21 August 1996).
26 www.optum.com/content/dam/optum/resources/productSheets/Clinformatics_for_Data_Mart.pdf (accessed April 9, 2019).
27 Hansen LG, Chang S. Health Research Data for the Real World: The MarketScan Databases. July 2011. Truven Health Analytics. http://truvenhealth.com/portals/0/assets/PH_11238_0612_TEMP_MarketScan_WP_FINAL.pdf (accessed April 9, 2019).
28 Reynolds K, Muntner P, Cheetham TC, *et al*. Primary non-adherence to bisphosphonates in an integrated healthcare setting. *Osteoporos Int.* 2013; **24**(9): 2509–17.
29 Medicaid Managed Care Enrollment and Program Characteristics, 2016. www.medicaid.gov/medicaid/managed-care/downloads/enrollment/2016-medicaid-managed-care-enrollment-report.pdf (accessed April 9, 2019).
30 CMS Fast Facts. January 2018. www.cms.gov/fastfacts/ (accessed April 9, 2019).
31 Research Data Assistance Center (ResDAC). Medicare Advantage (Part C) Encounter Data RIFs Final Version. https://www.resdac.org/cms-news/medicare-advantage-part-c-encounter-data-rifs-final-version (accessed July 29, 2019).
32 Byrd VL, Dodd AH. Assessing the Usability of Encounter Data for Enrollees in Comprehensive Managed Care 2010–2011. Washington, DC: Centers for Medicare & Medicaid Services, 2015. www.mathematica-mpr.com/our-publications-and-findings/publications/assessing-the-usability-of-encounter-data-for-enrollees-in-comprehensive-managed-care-2010-2011 (accessed April 9, 2019).
33 Hennessy S, Leonard CE, Bilker WB. Researchers and HIPAA. *Epidemiology* 2007; **18**(4): 518.
34 Hennessy S, Leonard CE, Freeman CP, *et al*. Validation of diagnostic codes for outpatient-originating sudden cardiac death and ventricular arrhythmia in Medicaid and Medicare claims data. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 555–62.
35 Ruttner L, Borck R, Nysenbaum J, Williams S. Guide to MAX Data. Medicaid Policy Brief #21. Mathematica Policy Research, 2015. www.mathematica-mpr.com/our-publications-and-findings/publications/guide-to-max-data (accessed April 9, 2019).
36 Transformed Medicaid Statistical Information System (T-MSIS). www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/index.html (accessed April 9, 2019).

37 Submitting Accurate and Complete Encounter Data (Managed Care). www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/?entry=43416 (accessed April 9, 2019).

38 Hennessy S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003; **12**(2): 103–11.

39 Kaiser Family Foundation. An Overview of Medicare. Issue Brief. November 2017. http://files.kff.org/attachment/issue-brief-an-overview-of-medicare (accessed April 9, 2019).

40 Kaiser Family Foundation. Medicare Part D in 2016 and Trends over Time. Publication #8915. http://files.kff.org/attachment/Report-Medicare-Part-D-in-2016-and-Trends-over-Time (accessed April 9, 2019).

41 Smith PM, Stock SR, McLeod CB, Koehoorn M, Marchand A, Mustard CA. Research opportunities using administrative databases and existing surveys for new knowledge in occupational health and safety in Canada, Quebec, Ontario and British Columbia. *Can J Public Health* 2010;**101**(Suppl 1):S46–52.

42 Downey W, Beck P, McNutt M, Stang M, Osei W, Nichol. Health databases in Saskatchewan. In: Strom BL, ed. *Pharmacoepidemiology*, 3rd edn. Chichester: John Wiley & Sons Ltd, 2000, pp. 325–45.

43 Abou Chakra CN, Moride Y, Greenfield B, *et al*. Validation of claims databases for the ascertainment of adverse events: from descriptive to predictive methodologies. *Pharmacoepidemiol. Drug Saf* 2009; **18**: S1–S273.

44 Population Data BC. www.popdata.bc.ca/data/ (accessed April 9, 2019).

45 Moride Y, Metge CJ. Canadian provincial databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 259–69.

46 Billioti de Gage S, Moride Y, Ducruet T, *et al*. Benzodiazepine use and risk of Alzheimer's disease: case-control study. *BMJ* 2014; **349**: g5205.

47 Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sorensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol* 2010; **106**(2): 86–94.

48 National Institute for Health and Welfare. Care Register for Health Care. https://thl.fi/en/web/thlfi-en/statistics/information-on-statistics/register-descriptions/care-register-for-health-care#data (accessed April 9, 2019).

49 Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015; **7**: 449–90.

50 Swedish National Patient Register. www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish (accessed April 9, 2019).

51 Guidelines for the Distribution of Data from the Norwegian Patient Register. https://helsedirektoratet.no/Documents/English/guidelines.pdf (accessed April 9, 2019).

52 Pottegard A, Schmidt SAJ, Wallach-Kildemoes H, Sorensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2017; **46**(3): 798.

53 Karlstad O, Zoega H, Furu K, *et al*. Use of drugs for ADHD among adults – a multinational study among 15.8 million adults in the Nordic countries. *Eur J Clin Pharmacol* 2016; **72**(12): 1507–14.

54 Yoshida K, Gruber S, Fireman BH, Toh S. Comparison of privacy-protecting analytic and data-sharing methods: a simulation study. *Pharmacoepidemiol Drug Saf* 2018; **27**(9): 1034–41.

55 Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol* 2010; **2**: 273–9.

56 Hallas J, Hellfritzsch M, Rix M, Olesen M, Reilev M, Pottegard A. Odense

Pharmacoepidemiological Database: a review of use and content. *Basic Clin Pharmacol Toxicol* 2017; **120**(5): 419–25.

57 Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT. Existing data sources for clinical epidemiology: the Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol* 2012; **4**: 303–13.

58 Wei L, Parkinson J, MacDonald TM. *The Tayside Medicines Monitoring unit (MEMO).* In: Strom BL, ed. Pharmacoepidemiology, 4th edn. Chichester: Wiley-Blackwell, 2005, pp. 323–36.

59 Bezin J, Duong M, Lassalle R, *et al*. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017; **26**(8): 954–62.

60 Coloma PM, Schuemie MJ, Trifiro G, *et al*. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011; **20**(1): 1–11.

61 Lai EC, Man KK, Chaiyakunapruk N, *et al*. Brief Report: databases in the Asia-Pacific Region: the potential for a distributed network approach. *Epidemiology* 2015; **26**(6): 815–20.

62 Page E, Kemp-Casey A, Korda R, Banks E. Using Australian Pharmaceutical Benefits Scheme data for pharmacoepidemiological research: challenges and approaches. *Public Health Res Pract* 2015; **25**(4): e2541546.

63 Department of Health. Sources of Epidemiological Data for Use in Generating Utilisation Estimates 2015. Canberra: Commonwealth of Australia. www.pbs.gov.au/info/industry/useful-resources/sources (accessed April 9, 2019).

64 Lalic S, Gisev N, Bell JS, Korhonen MJ, Ilomaki J. Predictors of persistent prescription opioid analgesic use among people without cancer in Australia. *Br J Clin Pharmacol* 2018; **84**(6): 1267–78.

65 Pratt N, Roughead EE, Ramsay E, Salter A, Ryan P. Risk of hospitalization for hip fracture and pneumonia associated with antipsychotic prescribing in the elderly: a self-controlled case-series analysis in an Australian health care claims database. *Drug Saf* 2011; **34**(7): 567–75.

66 Pratt NL, Ramsay EN, Kalisch Ellett LM, Nguyen TA, Barratt JD, Roughead EE. Association between use of multiple psychoactive medicines and hospitalization for falls: retrospective analysis of a large healthcare claim database. *Drug Saf* 2014; **37**(7): 529–35.

67 Roughead EE, Chan EW, Choi NK, *et al*. Variation in association between thiazolidinediones and heart failure across ethnic groups: retrospective analysis of large healthcare claims databases in six countries. *Drug Saf* 2015; **38**(9): 823–31.

68 Milea D, Azmi S, Reginald P, Verpillat P, Francois C. A review of accessibility of administrative healthcare databases in the Asia-Pacific region. *J Mark Access Health Policy* 2015; **3**.

69 Hsing AW, Ioannidis JP. Nationwide Population Science: lessons from the Taiwan National Health Insurance Research Database. *JAMA Intern Med* 2015; **175**(9): 1527–9.

70 Kim JA, Yoon S, Kim LY, Kim DS. Towards actualizing the value potential of Korea Health Insurance Review and Assessment (HIRA) data as a resource for health research: strengths, limitations, applications, and strategies for optimal use of HIRA data. *J Korean Med Sci* 2017; **32**(5): 718–28.

71 Park BJ, Sung JH, Park KD, Seo SW, Kim SW. *Report of the Evaluation for Validity of Discharged Diagnoses in Korean Health Insurance Database*. Seoul: Seoul National University, 2003, pp. 19–52.

72 Filion KB, Chateau D, Targownik LE, *et al*. Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis. *Gut* 2014; **63**(4): 552–8.

73 Ko CW, Dominitz JA, Green P, Kreuter W, Baldwin LM. Accuracy of Medicare claims for identifying findings and procedures performed during colonoscopy. *Gastrointest Endosc* 2011; **73**(3): 447–53 e441.

74 Levy AR, O'Brien BJ, Sellors C, Grootendorst P, Willison D. Coding accuracy of administrative drug claims in the Ontario Drug Benefit database. *Can J Clin Pharmacol* 2003; **10**(2): 67–71.

75 Wysowski DK, Baum C. The validity of Medicaid diagnoses of hip fracture. *Am J Public Health* 1993; **83**(5): 770.

76 Pottegard A, Christensen R, Houji A, *et al*. Primary non-adherence in general practice: a Danish register study. *Eur J Clin Pharmacol* 2014; **70**(6): 757–63.

77 Lloyd JT, Blackwell SA, Wei, II, Howell BL, Shrank WH. Validity of a claims-based diagnosis of obesity among Medicare beneficiaries. *Eval Health Prof* 2015; **38**(4): 508–17.

78 Crystal S, Akincigil A, Bilder S, Walkup JT. Studying prescription drug use and outcomes with medicaid claims data: strengths, limitations, and strategies. *Med Care* 2007; **45**(10 Supl 2): S58–65.

79 Stroup TS, Gerhard T, Crystal S, Huang C, Olfson M. Comparative effectiveness of clozapine and standard antipsychotic treatment in adults with schizophrenia. *Am J Psychiatry* 2016; **173**(2): 166–73.

80 Khaykin E, Eaton WW, Ford DE, Anthony CB, Daumit GL. Health insurance coverage among persons with schizophrenia in the United States. *Psychiatr Serv* 2010; **61**(8): 830–4.

81 Shin JY, Roughead EE, Park BJ, Pratt NL. Cardiovascular safety of methylphenidate among children and young people with attention-deficit/hyperactivity disorder (ADHD): nationwide self controlled case series study. *BMJ* 2016; **353**: i2550.

82 Lo Re V, Carbonari DM, Saine ME, *et al*. Postauthorization safety study of the DPP-4 inhibitor saxagliptin: a large-scale multinational family of cohort studies of five outcomes. *BMJ Open Diabetes Res Care* 2017; **5**(1): e000400.

83 Lo Re V 3rd, Haynes K, Ming EE, *et al*. Safety of saxagliptin: rationale for and design of a series of postmarketing observational studies. *Pharmacoepidemiol Drug Saf* 2012; **21**(11): 1202–15.

84 Moride Y, Lynd LD, Ducruet H, Li H, Tournier M, Greenfield B. Antidepressants and risk of suicide or self-harm in Canadian youth: a study involving common data models in Quebec and British Columbia. *Pharmacoepidemiol Drug Saf* 2014; **23**(S1): S10–S11.

85 Gupta N, Vujicic M, Blatz A. Opioid prescribing practices from 2010 through 2015 among dentists in the United States: what do claims data tell us? *J Am Dent Assoc* 2018; **149**(4): 237–45 e236.

86 Huybrechts KF, Gerhard T, Crystal S, *et al*. Differential risk of death in older residents in nursing homes prescribed specific antipsychotic drugs: population based cohort study. *BMJ* 2012; **344**: e977.

87 Schade Hansen C, Pottegard A, Ekelund U, *et al*. Association between QTc prolongation and mortality in patients with suspected poisoning in the emergency department: a transnational propensity score matched cohort study. *BMJ Open* 2018; **8**(7): e020036.

13

Electronic Health Record Databases

Daniel B. Horton[1], Harshvinder Bhullar[2], Lucy Carty[3], Francesca Cunningham[4], Alexis Ogdie[5], Janet Sultana[6], and Gianluca Trifirò[6]

[1] *Department of Pediatrics, Rutgers Robert Wood Johnson Medical School, Rutgers Center for Pharmacoepidemiology and Treatment Science, New Brunswick, NJ, USA*
[2] *Real-World Insights, IQVIA, London, UK*
[3] *Clinical Practice Research Datalink, Medicines and Healthcare products Regulatory Agency, London, UK*
[4] *Pharmacy Benefits Management, and Center for Medication Safety, US Department of Veterans Affairs, Hines, IL, USA*
[5] *University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*
[6] *Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of Messina, Messina, Italy*

Databases that contain health information can be divided into two broad categories: those that collect information for administrative purposes, such as filing claims for payment (administrative claims databases), and those that serve as the patient's medical record (electronic health record [EHR] databases), which physicians use to track health information about their patients. Administrative claims databases are maintained for billing or administrative purposes rather than for the actual provision of patient care. In contrast, data from EHR databases are thought more likely to be clinically accurate because they are collected for patient care, recorded by clinicians (versus coders), and reflect information that may not relate to billing. Unlike administrative databases, EHR databases are more likely to capture important health information about patients, such as symptoms of illness, historical data, family history, smoking and alcohol use, vital signs (e.g., body mass index [BMI]), and laboratory data [1]. Of note, we use the term *electronic health record* databases to encompass their use in the provision of clinical care (that is, as medical records) as well as the interoperability of these electronic systems across broad healthcare networks, serving various nonclinical functions, such as administration, billing, and research [2].

Despite their many advantages, EHR databases have certain limitations. Some EHR databases, like that of the Veterans Affairs (VA) and other US EHRs, may not capture diagnoses and treatments from out-of-system care. Other EHR databases, particularly European primary care databases, lack information from secondary care settings (e.g., hospitals and specialists), and linkage to secondary care datasets is not available for some databases. While EHR databases are thought to have greater clinical accuracy in recorded diagnoses, one cannot presume the validity of diagnostic codes without formal validation. EHR databases usually contain data on prescribed outpatient drugs, but many databases lack information on drug dispensing or inpatient medications. In addition, there may be a high proportion of

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Table 13.1 Overview of EHR databases.

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Country or countries (and region)	Italy (Caserta)	Spain	United Kingdom	Italy	Netherlands
Year data collection initiated	2000	2001	1987	1998	1989
Number of patients and amount of follow-up time*	0.6M patients, 3.4M person-years	7.9M patients; 49.7M person-years	30M patients (>15M research standard patients; 10M actively followed patients); >100M person-years	1.6M patients; 19.2M person-years	2.4M patients; >12.1M person-years [166]
Sex distribution*	Female: 0.31M (51.0%) Male: 0.28M (49.0%)	Female: 4.1M (52.3%) Male: 3.8M (47.7%)	Female: 49.3% Male: 50.7% [18]	Female: 0.83M (51.8%) Male: 0.77M (48.2%)	Female: 1.2M (51.2%) Male: 1.2M (48.8%)
Age distribution*	0–10: 0.01M (1.3%) 11–20: 0.05M (8.6%) 21–44: 0.22M (36.2%) 45–64: 0.18M (30.4%) 65–84: 0.11M (18.9%) ≥65: 0.03M (4.5%) Unknown: 0.03M (4.3%)	0–10: 1.2M (15.1%) 11–20: 0.72M (9.1%) 21–44: 2.9M (36.6%) 45–64: 1.7M (21.6%) 65–84: 1.2M (15.0%) ≥85: 0.20M (2.6%)	<18: 20.2% 18–64: 61.8% ≥65: 18.1% [18]	0–14: 0 (0%) 15–24: 0.08M (9.9%) 25–44: 0.48M (30.3%) 45–64: 0.54M (33.9%) 65–84: 0.35M (21.9%) ≥85: 0.06M (4.0%)	0–9: 0.26M (10.8%) 10–19: 0.27M (11.3%) 20–39: 0.64M (26.5%) 40–59: 0.65M (23.9%) 60–79: 0.46M (19.1%) ≥80: 0.13M (5.5%)
Race and ethnicity distribution	Not available	Mostly not available; some PCPs may have recorded race and ethnicity in free text	Available for 27% of patients in CPRD and 79% of inpatients in HES [167]. Known race/ethnicity distribution: White: 87% South Asian: 6% Black: 4% Mixed: 1% Other: 2%	Not available	Not available
Number of physicians or practices included*	300 GPs [34]	4910 GPs, 842 pediatricians	851 practices (actively contributing)	800 GPs [168]	Approximately 600 GPs from 200 practices [169]
Diagnostic coding system	ICD-9	ICD-9, ICPC	Read, ICD-10 (HES)	ICD-9	ICPC
Drug coding system	ATC, NDC	ATC	Gemscript	ATC	ATC

*Numbers updated through 2017.

IQVIA DA	Pedianet	SIDIAP	THIN	VA
Germany, France, United Kingdom	Italy	Spain (Catalonia)	United Kingdom	United States
1992	2000	2006	2002	1997
Germany: 34M patients (including 17.2M German specialty patients); 54.5M person-years France: 10.5M patients; 6.0M person-years [22] UK: 4.2M patients; 17.9M person-years	0.4M pediatric patients (0.2M actively followed); 1.8M person-years	5.6M patients; 5.7M person-years	18M patients (3.1M actively followed patients); >90M person-years	14.5M patients; 168M person-years
Germany Female: 19M (41%) Male: 15M (32%) Unknown: 12.5M (27%) France Female: 5.5M (52%) Male: 5M (47%) Unknown: 0.06M (1%)	Active patients: Females: 0.10M (48.2%) Males: 0.11M (51.8%)	Female: 2.9M (50.7%) Male: 2.8M (49.2%)	Female: 9.4M (52.2%) Male: 8.6M (47.8%)	Female: 1.9M (13.1%) Male: 12.5M (86.4%) Unknown: 0.07M (0.5%)
0–9: 1.5M (4.4%) 10–19: 2.6M (7.7%) 20–39: 7.9M (23.2%) 40–59: 9.5M (28.0%) 60–79: 8.2M (24.1%) ≥80: 4.3M (12.6%)	Active patients: 0–4: 0.04M (19.6%) 5–9: 0.06M (28.6) 10–16: 0.11M (51.8%)	0–10: 0.63M (11.2%) 11–20: 0.55M (9.7%) 21–44: 1.9M (33.3%) 45–64: 1.5M (27.1%) 65–84: 0.88M (15.6%) ≥85: 0.17M (3.1%)	Active patients: 0–10: 0.39M (12%) 11–20: 0.35M (11%) 21–44: 1M (32%) 45–64: 0.85M (27%) 65–84: 0.50M (16%) ≥85: 0.08M (2%)	<21: 0.24M (1.7%) 21–44: 2.4M (16.6%) 45–64: 3.5M (24.0%) 65–84: 5.3M (36.8%) ≥85: 2.4M (16.4%) Unknown: 0.65M (4.5%)
Not available	Not available	Not available	Not available	White 9.0M (62.4%) Black 1.9M (12.9%) Asian 0.24M (1.7%) Hispanic 0.48M (3.4%) Other 0.11M (0.8%) Unknown 2.8M (18.8%)
Germany: 2357 general practices, 2010 specialty practices France: 2091 practices UK: 218 practices [22]	300 family pediatricians	3414 GPs, 853 primary care pediatricians [170]	Over 700 practices	Healthcare professionals: 127 211 Physicians: 23 973
ICD-10, Read (UK)	ICD-9, ICD-10	ICD-10	Read, ICD-10 (HES)	ICD-9, ICD-10, CPT
ATC	ATC, NDC, Italian MINSAN codes	ATC, NDC	Gemscript	VA Drug Classification System [171], NDC

(*Continued*)

Table 13.1 (Continued)

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Software used	SANI.A.R.P.	Various, mainly OMI-AP [37]	Vision and EMIS	Millewin	Various
Quality checks, standards, and feedback to clinicians	Arianna conducts periodic data checks. GPs whose recorded data do not meet preestablished standards are excluded from research [34]	BIFAP performs extensive quality and validity checks of raw data from individual autonomous regions, including physician- and patient-level data. Patient data that are inconsistent or fail quality checks are excluded. BIFAP provides feedback to participating GPs and pediatricians by comparing their patients' registrations, disease characteristics, and prescribing indicators with those of other clinicians within BIFAP	CPRD performs permanent, ongoing quality checks of data from all practices. Patients with noncontiguous follow-up or poor data recording are excluded. Remaining patients are flagged as acceptable for use in research. See text for description of Up-to-Standard dates	IQVIA performs periodic quality checks based on coding accuracy, concordance between GP-specific and national prevalence of selected diseases, and mortality rates [168]. Data from GPs that do not meet set standards are excluded [25]	IPCI evaluates each GP practice for data quality based on several indicators. Data from practices below a preestablished quality threshold are excluded from research. GPs are not permitted to keep paper-based records to improve data quality [172]
Data access and approval	Arianna is available through collaboration with either the Local Health Unit of Caserta or academic institutions with data access, such as the University of Campania or the University of Messina. Researchers must first notify their local ethics committee before using the data. Full ethics committee evaluation is not needed (anonymized data, no direct patient interaction)	BIFAP is available to affiliated and other noncommercial researchers. Investigators must receive approval from the BIFAP scientific committee	CPRD (www.cprd.com) licenses online access to the database. Researchers can download CPRD data using a secure file transfer protocol. CPRD receives annual regulatory ethics approval to supply anonymized linked data for public health research. All research requests to access data held by CPRD are reviewed by the Independent Scientific Advisory Committee	Within Italy, LPD Italy (www.healthsearch.it/?lang=en) is available through collaboration with the Italian College of General Practitioners or with IQVIA (www.iqvia.com)	Access to IPCI (www.ipci.nl/Framework/Framework.php) is provided through collaboration with the Erasmus Medical Centre. Protocols for studies using IPCI data must be approved by the IPCI ethics committee [174]

IQVIA DA	**Pedianet**	**SIDIAP**	**THIN**	**VA**
Various	Junior Bit	e-CAPTM	Vision	Various
IQVIA checks all data for quality standards and plausibility. IQVIA gives all physicians monthly feedback reports showing their prescription patterns and those of colleagues within the IQVIA panel and within their specialty group. Data from DA Germany are also checked annually by the German Medical Association	Data quality in Pedianet is evaluated for every study conducted, either by a central database unit or by researchers. Quality checks include validation of ICD-9-based diagnoses in the clinical chart and free text	SIDIAP performs systematic quality checks to harmonize data and identify duplicate patients, logical errors, and implausible values and inconsistent units within laboratory data. SIDIAP initially used the Registry Quality Standard score to assess data quality [173], but this has been discontinued	IQVIA performs ongoing consistency and integrity checks on all THIN data. See text for description of Acceptable Mortality Reporting	All VA data are updated regularly and checked for quality. Drug data are updated daily with quality checks
IQVIA (www.iqvia.com) administers DA France or Germany with various options: researchers can buy the software and data with monthly updates, preprocessed datasets, or data analyzed directly by IQVIA. Approval for IQVIA DA France and Germany requires only local IRB approval	Pedianet data (pedianet.it/en/about) can be obtained by collaboration with Pedianet-affiliated epidemiologists	Researchers must pay a fee and sign an agreement to obtain data from SIDIAP. Commercial organizations may not use SIDIAP data directly but can contract the core SIDIAP research team to conduct studies with input from the scientific and ethical committee	IQVIA (www.iqvia.com) makes THIN data available in a few forms: a sublicense for the whole research-formatted dataset, a data subset, or a preprocessed dataset with some data manipulation. THIN studies require approval by the Scientific Review Committee of independent researchers. If additional information will be collected, ethics approval is required from the NHS Multi-Centre Research Ethics Committee	Access to VA data is limited to researchers employed by the VA or with VA appointments, and their collaborators. Approval by the local or central VA IRB is required

ATC, Anatomical Therapeutic Chemical; CPT, Current Procedural Terminology; GPs, general practitioners; HES, Hospital Episode Statistics; ICD, International Classification of Diseases; ICPC, Classification of Primary Care; M, million; NDC, National Drug Code; NHS, National Health Service; PCP, primary care professional.

Table 13.2 Selected variables in EHR databases available for epidemiologic research.

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Healthcare professional demographics	Arianna provides physicians' age, sex, years since graduation	BIFAP provides geographic area of GP but not GPs' demographic data. One cannot determine whether nurse or doctor entered data	CPRD reports if nurse or doctor entered data	One can identify GPs' geographic region but not demographics	GPs' demographics are not available
Types of physicians	GPs	Mainly GPs but also other members of the primary care team, such as pediatricians and nurses	GPs	GPs	GPs
Practice and patient demographics	Practice: Location Patient: DOB, sex, healthcare exemption (based on salary and disability)	Practice: Number of patients registered with GP; number of persons registered in practice available upon request Patient: DOB, sex	Practice: Region, practice size, practice-level SES (Index of Multiple Deprivation and Townsend scores, ~60%), date of last registration, Up-to-Standard date (see text) Patient: YOB for adults, month and YOB for children; sex, ethnicity (~25% recorded; also available via census data), census-based socioeconomic class; patient status (active, died, transferred out)	Practice: Location Patient: DOB, sex healthcare exemption (based on salary and disability)	Practice: Number of employees may be available for some practices Patient: DOB, sex
Vital signs and social history	Height, weight, BMI, smoking, alcohol use available for 25% persons aged ≥65 (2013–present); BP available for some patients (2016–present)	Weight, BMI, BP, smoking, and alcohol consumption	Height, weight, BP, and BMI recorded but may be biased towards patients with a more relevant need for these measurements; smoking (83–93%) [175,176], obesity (61–79%) [175–177], alcohol (~80%) [175,178]	BMI [179], BP [172], smoking, alcohol intake [180]	BP, weight, BMI, and smoking [181] available but recorded only when GPs consider them relevant

IQVIA DA	Pedianet	SIDIAP	THIN	VA
IQVIA DA provides physicians' age, sex, and years in practice	Pedianet provides pediatricians' age, sex, and city of clinic. More detailed information (e.g., years since graduation) is available by request	SIDIAP provides age, sex, type of primary care professional, and performance indicators (quality of care, quality of prescriptions, and quality of diagnosis)	THIN reports if nurse or doctor entered data	One can identify the doctor, nurse, or pharmacist who entered prescription data
Mainly GPs; IQVIA DA Germany and France also include specialists (e.g., cardiologists, dermatologists)	Family pediatricians	Health professionals working in primary care: GPs, pediatricians, dentists, nurses, midwives	GPs	Mainly PCPs but also physician specialists (e.g., cardiologists) and other clinicians (e.g., nurse practitioners, clinical pharmacy specialists)
Practice: Region, community size, patients per practice, number of doctors, number of employees, type (e.g., GP vs specialty) Patient: Age, sex, health insurance status (e.g., private, statutory), medical insurance company, region, town size (>100 000 vs <100 000)	Practice: Region, patients per practice Patient: YOB, age, sex, region of residence, nationality, information about parents (e.g., nationality, habits, blood group, mother's educational level, socioeconomic level)	Practice: Location, urban/rural, number of patients, deprivation index (MEDEA) Patient: DOB, sex, country of origin	Practice Region, number of patients, computerization date, Vision date, Acceptable Mortality Reporting (see text) Patient: YOB for adults, month and YOB for children; patient-level, location-based socioeconomic status (Townsend deprivation scores, 95% recording), region, ethnicity, (20% recording), patient status (active, died, transferred out)	Practice: Region, facility, type of facility (medical center clinics vs community-based outpatient clinics), facility's level of complexity Patient: DOB, sex, race, ethnicity, zip code
BMI (~40% [182]); smoking and alcohol recording unknown	Gestational age, birth weight, birth height, neonatal jaundice; growth measurements (e.g., height, weight); parental smoking	BP, BMI, smoking, alcohol intake, Framingham score. Pediatric screening data (height, weight, head circumference, pubertal development)	Height, weight, BP, and BMI recorded but may be biased towards patients with a more relevant need for these measurements; smoking (86–94%) [86,183–185], obesity (73–83%) [185], alcohol intake (75–85%) [185]	BP, HR, height, weight, SES, education, marital status, smoking history (>90%)

(*Continued*)

Table 13.2 (Continued)

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Referrals, procedures, results of investigations	Laboratory test results for ~25%; linkage to hospital discharge data, referral data, and orders for diagnostic tests	PCPs' referrals to specialists and hospitals; results from referrals may be recorded in coded fields or as free text; self-referrals (less common) not available	Detailed information on referrals, procedures, and laboratory tests available for approximately 75% of all patients through linkage to HES	Referral data and orders for diagnostic tests available for all patients	Often not available; letters from hospitals to GPs with test results may be available for some practices; test results may be manually recorded in free-text notes
Type of drug data	Drugs prescribed in community setting; drug dispensing by linkage to claims	Drugs prescribed and dispensed in community setting; vaccine data available	Drugs prescribed in primary care; some OTC drug data available (see text); vaccine data available	Drugs prescribed in community setting; vaccine data available	Drugs prescribed in community setting; vaccine data available
Available drug information	Drug ATC code, NDC (with brand, formulation, units), indication for use	Drug name, active substance, number of prescribed packages, duration, prescribed daily dose, strength, indication for use	Drug name, route, strength, frequency, duration; immunizations including batch; cost of therapy upon request	Drug name, route, dose, frequency, duration, cost of therapy	Drug name, quantity, strength, dose [187]
Health care utilization	GP visits; hospital discharge letters, referrals to specialists, admission to ED available by linking with claims data	GP visits; referrals by GP to secondary care and ED; hospital admissions available if patients referred to GPs after discharge	GP visits, hospitalizations, and consultant visits; links to HES provide detailed ward-level resource utilization (England only)	GP visits, hospital discharge letters, referrals to specialists	GP visits; other data generally not available unless hospital discharge letters sent to GP
Identification of pregnancy and families	ICD-9 codes for pregnancy or birth by linkage to claims data; cannot identify families	ICD-9/ICPC codes for pregnancy; cannot identify families	Pregnancy and pregnancy outcomes, family/ household identification number; mother–baby link via family/ household number and algorithm	Not available	Some birth-related data available through hospital discharge letters; cannot identify families

IQVIA DA	Pedianet	SIDIAP	THIN	VA
HbA1c, blood glucose, cholesterol, LDL, HDL available; other test results variably available but can be requested from paper files	Apgar scores, laboratory and imaging tests ordered and reasons for request; test results not always available	Laboratory tests (date, results), diagnostic and imaging referrals; spirometry; referrals for therapeutic procedures; referrals to secondary and tertiary care [186] (date, reason of referral [ICD-10], specialty referred)	Electronic referrals available; older referrals may be in paper files; most outpatient laboratory results available	Provider referrals for specialists available; all laboratory results available but must be standardized
Drugs prescribed	Drugs prescribed and dispensed in community setting, including drugs not covered by healthcare system; inpatient drug data available if reported to pediatrician; noncompulsory vaccine data available, remaining vaccine data identified via linked claims	Drugs prescribed and dispensed in community setting for drugs covered by the national healthcare system; vaccine data available	Drugs prescribed in primary care; vaccine data available	Drugs prescribed and dispensed in outpatient and inpatient settings; vaccine data available
Drug name, route, dose, frequency, duration, cost of therapy	Drug name, ATC code, indication for use, Italian MINSAN code, NDC (with brand, formulation, units), number of prescribed packages, dose (not available for 30%)	ATC code, NDC, indication for use, profession of prescriber; prescribing data only: start and end date, drug units per day; dispensing data only: units per package, number of packages per month, month of drug dispensation	Drug name, route, strength, frequency, duration; immunizations including batch; linkage available to cost of therapy	Drug name, route, strength, dose, frequency, quantity, duration; cost of therapy
GP visits, hospitalizations, sick leave	Pediatrician visits, ED or hospital admission if referred by pediatrician	PCP visits, referrals to secondary and tertiary care, sick leave (date, length, ICD-10), hospital discharge	GP visits, hospitalizations entered by GP, sick leave (if issued by GP); links to HES provides detailed ward-level resource utilization (England only)	Outpatient visits, ED visits, hospitalization (including medical surgical, and intensive care units), community living center (VA nursing home)
Pregnancy variable, gynecologist records; family data incomplete	May identify siblings	Pregnancy and pregnancy outcomes; mother–baby link available	Pregnancy and pregnancy outcomes; mother–baby link via family/household number and algorithm	ICD-9/ICD-10 codes for pregnancy

(*Continued*)

Table 13.2 (Continued)

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Identification of death and cause of death	Date of death by linkage to claim data	Date of death; cause of death not available consistently	Date and cause of death available via CPRD data and linkage to Office for National Statistics	Date of death	Date of death; cause of death available via free text
Additional data, such as consult records, free text, or paper files	No free text or letters available	Anonymized free-text notes by GPs are available	Hospital discharge summaries, consultant letters; no free text available	No free text or letters available	Free text available on request
Questionnaires and investigator-initiated outcome validation	Not possible to administer questionnaires	Questionnaires can be given to GPs	Questionnaires can be given to GPs and patients; response rates from three recent studies were ~90% [188] (and CPRD internal data)	Not possible to administer questionnaires	Possible to administer questionnaires but response rates usually low
Settings and types of missing data	Inpatient data (except via discharge forms with main diagnoses), laboratory results for 75%, OTC drugs, vaccines	Inpatient data, OTC drugs (few OTC drugs in Spanish national healthcare system)	Prescriptions in secondary care, OTC drugs (exceptions in text), drug dispensing, adherence	Inpatient data, OTC drugs, drug dispensing, pediatric clinical and prescribing data (any setting)	Inpatient and specialist data, OTC drugs, drug dispensing; linkage available to Dutch PHARMO database with dispensing data

IQVIA DA	Pedianet	SIDIAP	THIN	VA
Date and cause of death seldom recorded	Date and cause of death	Date of death	Death date, sometimes cause of death; death certificates may be accessed for a fee if ethics approval is obtained	Date of death
No free text available	Free text available on request	Hospital discharge for 30% of the SIDIAP patients; other data available by request	Hospital discharge summaries, consultant letters; no free text available	Additional data including consult records and free text available by chart review
Questionnaires available upon request	Patients and families can be contacted for structured or unstructured interviews by phone calls from participating pediatricians	Questionnaires can be given to sample of GPs	Questionnaires can be given to GPs and patients; response rates to paper questionnaires ~90% [51] (and THIN internal data)	Charts may be reviewed for validation
Secondary care records, vaccine data, linkage between patients seen in both primary care and specialty clinics	Inpatient data not available for 60%; OTC drug data (unless reported to pediatrician); adult health data	Inpatient data not available except admission/discharge data for hospitals of the Catalan Health Institute; OTC drugs, indication for drug use, drugs not covered by national health system	Prescriptions in secondary care, OTC drugs (exceptions in text), drug dispensing, adherence	Encounter and drug data from healthcare facilities outside VHA, including for patients taken to nearby hospitals for acute events (e.g., stroke); some inpatient medications housed in floor stock for acute care

BP, blood pressure; DOB, date of birth; ED, emergency department; GPs, general practitioners; HES, Hospital Episode Statistics; ICD, International Classification of Diseases; ICPC, International Classification of Primary Care; NDC, National Drug Code; OTC, over the counter; PCP, primary care professional; SES, socioeconomic status; VHA, Veterans Health Administration; YOB, year of birth.

missing data for some variables of interest, such as disease severity, smoking history, body mass index, and patients' race or occupation.

In this chapter, we focus on selected primary care EHR databases from Europe and a national EHR database for veterans from the United States. Several European databases are available for licensing by investigators in government, academia, and industry, while access to certain European databases and to VA data requires collaboration with affiliated researchers. These databases have been used by epidemiologists, and in particular pharmacoepidemiologists, resulting in thousands of published studies. While there are many similarities among the databases, there are also important differences, which we describe in more detail below (see also Tables 13.1 and 13.2). Of note, we do not discuss other outpatient EHR databases that may have been used in pharmacoepidemiologic research but are less widely utilized or less representative of broader source populations. Claims databases from various countries are covered in Chapter 12, and inpatient EHR databases will be discussed at length in Chapter 14.

Description

Europe and the United Kingdom

Overview of Healthcare Systems and Populations

France, Italy [3], Spain [4], and the UK have universal, government-funded healthcare systems. Germany and the Netherlands require all persons to have medical insurance to cover healthcare [5]. In several of these and other European countries, general practitioners (GPs) act as gatekeepers for medical care. In Italy and Spain, family pediatricians function similarly as gatekeepers of most children's healthcare. Except in certain European countries (e.g., France, Germany), practically the entire population have primary care professionals (GPs or family pediatricians), and the vast majority of these clinicians have EHRs. Where GPs and pediatricians act as gatekeepers of the health system, they not only provide general (primary) medical care but are also involved in or informed of nearly all medical events involving their patients, including referrals to specialists, admission to emergency departments or hospitals, and prescribing of medicines recommended by consulting specialists. Thus, European primary care-based EHR databases capture most of their patients' health information.

Of note, the Italian and Spanish healthcare systems are strongly decentralized [6]. Healthcare services in these countries are managed and provided at the regional level, and their respective EHR databases reflect this regionalization. Notably, while both France and Germany have universal healthcare systems, patients often have additional private insurance, and GPs have less of a gatekeeper role than in other countries. These more open healthcare systems, therefore, make the French and German EHR databases less complete records of patients' health information [7–10].

Overview of Databases

The UK was the setting of the first European EHR database, Clinical Practice Research Datalink® (CPRD®), (previously known as Value Added Medical Products [VAMP] database and then the General Practice Research Database® [GPRD®]). CPRD was established in 1987 as a tool for conducting public health research. The Dutch IPCI database followed shortly thereafter in 1989. Since then, multiple other European health record databases were developed and used for research purposes (see below and Table 13.1).

Clinical Practice Research Datalink is a research service of the UK government, supported by the Medicines and Healthcare products Regulatory Agency (MHRA) and National Institute of Health Research. The Health Improvement Network® (THIN®) was set up in 2002 as a collaboration between software and database companies (respectively, Cegedim and

Epic Database Research Company Ltd, now part of IQVIA; THIN is a Cegedim database). CPRD and THIN collect similar health information from approximately 3–5 million individuals per year seen by GPs in the UK, representing 5–8% of the population [11–19]. The same practices may contribute to both CPRD and THIN, but the proportion of overlap changes over time as new practices join or leave each database.

In a study validating well-established drug–outcome associations from the literature, findings were similar in CPRD practices and non-CPRD practices within THIN [16]. As an example of the overlap of these two databases, one study identified over 60% of individuals initiating a particular drug in both THIN and CPRD [20]. Increasingly, pharmacoepidemiologists are combining information from both databases to increase sample size and improve statistical power and generalizability. Because CPRD and THIN are not mutually exclusive, merging data requires identification and singular inclusion of practices contributing to both databases in a given year. Investigators have developed an algorithm for identifying overlapping practices while maintaining anonymity based on total numbers of patients per practice stratified by gender and birth year [21].

The IQVIA Disease Analyzer databases (DA, previously known as Mediplus®) were set up in France, Germany, and the UK. Because the DA UK database is no longer available, it will not be discussed in detail, but additional information may be found in Tables 13.1 and 13.2. The IQVIA DA databases include anonymized patient records from primary care practices as well as some office-based specialists, including cardiologists, dermatologists, diabetes specialists, gynecologists, neurologists, orthopedists, otolaryngologists, pediatricians, psychiatrists, and urologists [22,23]. Patients who see both general practitioners and specialists have different identity codes in the databases to preserve patient confidentiality, making it challenging to track patients across different settings of care. With nearly 30 million patients (5–7% of the total population), DA Germany is larger than DA France, which has over 10 million patients (16% of the total population).

In Italy, the Health Search Longitudinal Patient Database (LPD Italy from IQVIA) contains data on 1.6 million people from GPs across the nation (2.6% of the population), making it the country's largest EHR database [24–29]. Founded in 1998 by the Italian College of General Practitioners, LPD Italy is now owned by IQVIA. The Arianna database contains EHR data on approximately 600 000 people (60% of inhabitants) in a region of southern Italy [30–34]. The Arianna database is the only Italian EHR database that systematically links to several administrative claims databases and includes drug dispensing and hospital discharge data (see Chapter 12 for other data sources combining claims and EHR data). Of note, Arianna data may be linked to comprehensive geriatric assessments (systematic, multidimensional evaluations of health status covering cognitive and physical function, mobility, disability, social support, etc.) for almost three-quarters of the local elderly population (90 000 since 2014), making it a valuable resource for geriatric research [35]. On the other side of the age spectrum, Pedianet contains data on over 400 000 children throughout Italy since 2000, over half of whom are being actively followed (see Chapter 22 for other data resources for pediatric pharmacoepidemiologic research).

In Spain, Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP) and Sistema de Información para el Desarrollo de la Investigación en Atención Primaria (SIDIAP) are primary care EHR databases that differ in catchment area. Founded in 2000 by the Spanish Agency on Medicines and Medical Devices, BIFAP contains data on 7.9 million individuals from nine of 17 autonomous communities (17% of the Spanish population) [36,37]. SIDIAP was founded in 2006 by the Catalan Institute of Health and the Primary Care Research Institute Jordi Gol. SIDIAP contains data on almost 5.6 million persons from a single

autonomous community, Catalonia (85% of the Catalonian population).

General practitioners across the Netherlands contribute data to Integrated Primary Care Information (IPCI, previously known as Interdisciplinary Processing of Clinical Information). As of 2018, 2.4 million individuals (14% of the Dutch population) are registered in IPCI.

Data Collection and Structure

Practitioners use electronic health records to document a wide range of clinical information about their patients. These data are then electronically extracted for research purposes by specialized software (see Table 13.1), examined for completeness and accuracy by database administrators, and uploaded in anonymized form into the database. The frequency of data extraction and transmission varies among databases. Some EHR databases receive frequent data updates (e.g., real-time contributions in Pedianet, daily in CPRD, three times per week in THIN, monthly in Arianna). Other databases receive new clinical data just 1–2 times per year (e.g., BIFAP, IPCI, LPD Italy). Nonetheless, all of these databases continue to accrue new information over time, adding data on new patients entering the healthcare system and updating data on existing patients who are followed longitudinally. There is substantial heterogeneity in how information is extracted for research across databases, and sometimes within databases, reflecting the diversity of software and healthcare systems. For example, in Spain, some autonomous regions send data to BIFAP directly, whereas other regions use local software programs for data extraction.

Primary care EHR databases generally contain a minimum set of clinical information, including data on patient demographics, medical diagnoses, and drug prescriptions. Table 13.2 contains detailed lists of data collected in each database and highlights differences among them.

European EHR databases use a variety of standardized coding systems to record diagnoses: Read codes (THIN and CPRD); International Classification of Diseases, 9th edition (ICD-9) (Arianna, one autonomous region in BIFAP, LPD Italy, and Pedianet); ICD-10 (IQVIA DA, SIDIAP, part of Pedianet, Hospital Episode Statistics [HES] data linked with CPRD and THIN); and the International Classification of Primary Care (ICPC) (IPCI, most autonomous regions in BIFAP).

The European EHR databases also vary in the ways they record drug data. CPRD and THIN employ British National Formulary (BNF) codes through the Gemscript system. European databases outside the UK record medications using Anatomical Therapeutic Chemical (ATC) classification codes, although many countries have national drug codes contained within a governmental formulary or similar compendium. All European EHR databases discussed in this chapter contain data on prescribed medications. Arianna, BIFAP, Pedianet, and SIDIAP also contain drug dispensing data. Additionally, Arianna, BIFAP, Pedianet, and (for roughly half of drugs) DA databases specify the indications for drugs.

All the above databases are representative of their respective source populations in terms of the distribution of age and sex and the prevalence of most diseases and prescribed drugs [18,19,22]. However, because data are collected for clinical and not research purposes, the reported frequency of certain diseases may vary across databases depending on local or disease-specific patterns of clinical care [38,39]. BIFAP, CPRD, DA France and Germany, IPCI, LPD Italy, and THIN include most regions of their respective countries. However, the distribution of patients across regions in these databases may not reflect the actual populations of those regions [1,15,17,40]. Similarly, the spectrum of socioeconomic status found in the databases may differ from the country as a whole [40].

Vital signs (e.g., blood pressure, height, weight, BMI) and laboratory test results are available in Arianna, BIFAP [41], CPRD [42], DA [43], IPCI, Pedianet, SIDIAP [44–46], and THIN [47] to varying degrees (see Table 13.2). For example, in both THIN and CPRD, laboratory data from after approximately 2000 is better recorded, but some older laboratory tests may not be available electronically if they were received by GPs on paper. Furthermore, height and weight are recorded for most adults in the UK but may be missing for many children (see Incompleteness of Clinical Data). In DA Germany, HbA1c for diabetic patients is nearly complete, but many other laboratory values are not recorded. Laboratory test results are also available in Arianna (~25% of patients), BIFAP, Pedianet (with indications for testing), and SIDIAP.

Hospitalizations, referrals, and the resulting consultation letters are recorded to varying degrees in European EHR databases. In the UK, discharge summaries and other hospital and consultant letters are sent to the GP, although these identified paper documents are not directly available to researchers. Linkage of THIN and CPRD to HES data allows researchers to access additional details from hospitalizations, such as diagnosis codes on admission and discharge and length of hospital stay. Referrals to other care settings are captured in both CPRD and THIN, and data from outside consultations may be obtained by linkage to HES data. Details on available data for hospitalizations and referrals in other databases are listed in Table 13.2.

Data from social history, including smoking and alcohol usage, are available to varying extents in most EHR databases (see Data Quality: Accuracy and Completeness, and Table 13.2). Substance exposure information is less consistently recorded in DA databases and IPCI. Pedianet contains information on parental smoking habits. Certain components of social history, such as occupation, are not routinely recorded in some databases [48].

In most of the European EHR databases described, most data are entered using structured (coded) fields rather than free text [49–52]. In contrast, BIFAP and IPCI contain large volumes of unstructured data. These and other databases make information from anonymized free text entries available to researchers (see Table 13.2). These free text data can be used to identify and validate outcomes and supplement available data in structured fields. Some practitioners may still keep paper record files, which could include precomputerization records, hospital discharge paperwork, or letters from specialists. Multiple databases, including THIN and CPRD, have additional fee-based data services that will obtain and anonymize paper-based data from GPs (see Table 13.2). To maximize the quality of the electronic record, IPCI does not permit participating GPs to keep paper records, but Dutch GPs record extensive free text notes, which are available to researchers [29].

Aside from the availability of unstructured data, several databases, including BIFAP, CPRD, DA Germany, Pedianet, SIDIAP, and THIN, allow researchers to administer questionnaires to clinicians or patients [51,53–55]. Like free text data, such questionnaires can be used to validate existing data or provide additional information that is not otherwise available in the database. Moreover, unlike free text entries, such questionnaires can be tailored in their content and administration based on the specific research question and population of interest [51,56]. Researchers must pay fees to administer supplementary questionnaires, a portion of which participants receive to complete the questionnaires. Of note, investigator-initiated surveys are also permitted in IPCI, but anecdotally response rates among GPs tend to be low.

Data Quality: Accuracy and Completeness

Data quality checks are performed by the European EHR databases at regular intervals on three levels: (1) practitioner recording, (2) data

extraction, and (3) maintenance of the database (see Table 13.1 for details on database-specific measures of quality assurance). When data are uploaded or extracted from the health records, the company processing the data performs additional quality checks to make sure the data have been correctly uploaded or extracted. Subsequent updates to the databases are verified for accuracy [1,40,57]. All databases undergo routine updating of the software used to collect, check, transfer, and present data.

In the UK, national quality improvement initiatives as well as advances in software have increased overall capture and accuracy of data. The national UK initiative the Quality and Outcomes Framework (QOF), a pay-for-performance program, was instituted in 2004 to improve performance using 146 quality indicators for 10 chronic diseases [58]. The QOF measures increased GP input in the EHR, leading to more complete data recording, especially for the targeted medical conditions [59,60]. Even after certain financial incentives were removed, performance for many quality indicators persisted across UK practices [61]. However, it is unclear whether reporting also improved for other, nonspecified quality indicators or diseases [62]. Some researchers have reported that the QOF has not contributed to decreases in mortality, better care coordination, or better patient experiences in the UK [63]. In its current form, therefore, this program's future is uncertain [64].

The QOF is only one of multiple quality improvement strategies implemented in the UK since the 1990s [62,65]. In coordination with specific databases, GPs receive training in the use of their software and regular evaluation of their data recording and prescribing behavior. GPs contributing to THIN or CPRD receive feedback reports with tips on improving recording and, in some cases, a summary of their prescribing habits relative to similar practices and other GPs in the UK. Other database-directed quality measures include audits of newly added practices and comparison of acquired data to national databases (e.g., mortality, hospitalizations, cancer, and cardiovascular registries) [22,66]. Finally, as incentives to take part in research studies, GPs and practices contributing to CPRD or THIN may receive income through questionnaires or participation in interventional clinical studies.

In CPRD, only data from practices that meet quality standards (~90% of practices) are provided for research. The Up-to-Standard (UTS) date is a practice-based quality marker corresponding to when a practice in CPRD is considered to have continuous and complete recording of data [17]. The UTS date is based on two parameters: the presence of gaps in the data stream and the existence of an appropriate rate of recorded deaths at the practice. For THIN, IQVIA employs a quality measure known as acceptable mortality reporting (AMR), denoting the year in which mortality reporting was deemed complete for each practice [40,53]. Other European EHR databases have their own standards for ensuring quality and completeness (see Table 13.1 – Quality checks, standards, and feedback for details).

With regard to specific variables, completeness of data varies among databases (see Incompleteness of Clinical Data and Table 13.2). Pregnancy, family structure, mortality, and cause of death are variably recorded and may be difficult to ascertain. Family structure may require the use of coding algorithms [66–74]. CPRD offers a probabilistic mother–baby link algorithm, which identifies likely mother–baby pairs based on an anonymized family number, maternity information from the mother's primary care record, and the month of birth of newly registered babies. CPRD may also be linked to mortality records from the Office for National Statistics to improve death estimates and confirm cause of death [75–77]. Researchers may also use algorithms for data in THIN to link family members or determine cause-specific mortality [78,79]. Risk factors such as smoking

and obesity may have gaps. The introduction of QOF measures led to a substantial increase in recording of these and other variables within CPRD and THIN [80–85]. For example, recording of data on adult patients' smoking habits rose from 75% before 2004 to nearly 90% by 2007 [86]. By 2005, 99% of patients with diabetes had a reported HbA1c value in the past 15 months compared to 87% in 1998.

As with other data sources, investigators need to consider the local context when interpreting dates in European EHR databases. Dates in the medical files may reflect dates of data entry or dates on which observations were made. In the case of new registrations in general practice, dates may reflect entry of data obtained from previous practitioners or from previous recording systems.

Data Access for Researchers

Research performed using European EHR databases must first be reviewed by the home institution's institutional review board (IRB) and the ethics board for the respective database. Given researchers' inability to identify individual patients in anonymized databases, such studies often meet the criteria for IRB exemption. Investigators must usually receive approval from the ethics board of the respective database before conducting their research. Requirements for approval change over time, so investigators should check with the data vendor about approval requirements prior to starting a study. Companies may also require completion of a data use agreement before initiation of a study. See Table 13.1 (Data access and approval) for more details.

United States: Department of Veterans Affairs Healthcare

Overview of Healthcare System and Population

The Department of Veterans Affairs (VA) was established in 1930 as the Veterans Administration based on congressional approval to "consolidate and coordinate Government activities affecting war veterans" [87,88]. The VA's Veterans Health Administration (VHA) is one of the largest integrated healthcare systems in the United States, providing medical, surgical, and rehabilitative care to a diverse group of military veterans as well as active duty reservists and National Guard. In contrast to the general US population, the VA population consists of predominantly older men (87% male, 47% over age 65 as of 2017) who often have multiple chronic medical or psychiatric conditions. The female population in the VA has increased over the last several years and represents a younger cohort of veterans. In 2016, the VA healthcare system consisted of 18 regional integrated networks encompassing 145 hospitals and medical centers, over 1200 ambulatory care, mobile, independent, and community-based outpatient clinics, and 132 community living centers (VA nursing homes) [89,90].

The VHA is primarily a direct provider of healthcare services, funded by the US government. While veterans receiving healthcare are not required to pay premiums for coverage, some are charged co-payments for certain medical services and outpatient prescriptions [88]. The vast majority of medications within the VHA are prescribed by VA clinicians and dispensed by VA pharmacies. To facilitate access to care, veterans may also see and receive medications from certain authorized private providers outside the VA [91]. Prescriptions from this program (<1% of all VA prescriptions as of 2017) are similarly dispensed and recorded within the VHA. Of note, dual-care veterans with Medicare coverage (which covers virtually all US citizens age 65 years and older) may receive medications through both the VA and the Medicare Part D Plan. This arrangement may result in duplicate or excessive drug usage for some patients – a clinically important situation that might be overlooked by clinicians and researchers alike who do not jointly consider both sources of medical care [91–93]. Because the VHA is not a

closed medical system, veterans may receive out-of-network care, compromising one's ability to study certain outcomes (see Incompleteness of Clinical Data).

In 2017, approximately 6.3 million veterans (more than one-third of US veterans) were treated in the VA healthcare system, with over 5 million receiving prescriptions (see Table 13.1 for details) [89].

Overview of Database

The VA database contains demographic, clinical, and administrative data from 1997 to the present along with prescription data since 1999 and laboratory data starting in 2002. This database contains health information on 14.5 million patients; in 2017, the number of veterans in the VA database was about 2% of the US population and 32% of US veterans [89].

Data Collection and Structure

Local VA medical centers record and store clinical and administrative data within the Veterans Health Information Systems Technology and Architecture (VistA) system. The National Patient Care Database (NPCD) contains inpatient data and (through 2018) outpatient data extracted, organized, and integrated from VistA. These data include demographics, diagnoses (ICD-9 and ICD-10 codes), clinic visits, admissions, discharges, transfers, prescription orders, laboratory, surgical procedures, provider specialty, and administrative services [94,95]. The VA's Corporate Data Warehouse (CDW), a primarily operational database, contains raw clinical, medication and administrative data through VistA [96,97]. While less structured than the NPCD, the CDW contains similar information as well as additional data, including vital signs, radiology results, free text notes, consults, and health factors (e.g., smoking status). Since 2019, the CDW has been the primary source of outpatient VA data.

These data may be used for quality improvement efforts and research. The VA Vital Status File contains some demographics (birth date, gender) and death data from multiple sources [98]. Death data are cross-checked monthly with the Social Security Administration Death Master File [98]. In addition, VA has several disease-specific registries that are used for patient care and research, including cancer, diabetes, severe mental illness, amyotrophic lateral sclerosis, rheumatoid arthritis, and the human immunodeficiency virus and hepatitis C clinical case registry.

Unlike most EHR databases, the VA database contains information on both prescribing and dispensing of drugs in both outpatient and inpatient settings. Pharmacy data systems record outpatient and inpatient drug dispensing in the CDW and Pharmacy Benefit Management (PBM) databases [99,100]. The PBM database also contains information on dispensing of nonprescription medications and specific medical supplies. The majority (85%) of outpatient medications are dispensed via the VA's consolidated mail-order pharmacies. While the PBM database records the dispensing of inpatient drugs, the Bar Code Medication Administration (BCMA) database contains records of administrations of medications to inpatients [101]. The VHA also maintains its own adverse drug event reporting system (also see Chapter 10); as of 2018, this system contains over 500 000 reports related to drugs or vaccines.

Investigators may also extract data for research directly from the EHR. Such primary data collection can facilitate, for example, the validation or ascertainment of outcomes in unstructured portions of the EHR, such as reports (e.g., biopsies), text-embedded test results (e.g., cardiac ejection fraction from an echocardiogram), and free text notes. Researchers may extract data through manual chart review in the local web-based EHR portal or via natural language processing programs [102,103]. In addition to EHR data, surveys of veterans and clinicians permit access to additional information [104].

Data Quality: Accuracy and Completeness

The CDW, a nontransformed mirror of the medical record, is updated nightly. Updates of the Vital Status File occur monthly. The accuracy and completeness of data reflect the EHR and beneficiary claims they source (see Tables 13.1 and 13.2 for more details). For instance, demographic race and ethnicity data can have up to 20% missing data in certain years [105].

The PBM database undergoes daily quality assurance processes to ensure completeness and accuracy [99]. As with medications obtained through Medicaid (see Chapter 12), low or nil co-payments produce strong financial incentives for veterans to obtain outpatient prescriptions through the VA.

Data Access for Researchers

Researchers may obtain data through any of the aforementioned VA data systems (see Data Collection and Structure) following approval by the local or central VA IRB. Access to VA data is limited to VA-affiliated researchers and their collaborators.

Strengths

Population-Based Data, Sample Size, and Length of Follow-up

Population-based studies draw subjects from the greater population to produce a derived sample that reflects the source [106]. Many European EHR databases allow researchers to use population-based study designs, minimizing selection bias and improving the validity and generalizability of epidemiologic studies. Although patients can opt out of having their information used, few do so.

Population-based data sources are ideal for nested case–control studies, in which all cases (e.g., individuals with the outcome) or a representative subsample are ascertained in a precisely defined population, and unaffected controls are sampled randomly from the same source population at the time when cases develop the outcome (incidence density sampling) [107]. Similarly, population-based data allow for the design of cohort studies, given the availability of prospective data with long follow-up periods. The large numbers of patients with longitudinal follow-up (see Table 13.1) may allow for sufficient statistical power to study rare exposures, diseases, and outcomes. These large, population-representative databases are excellent settings for a wide variety of methodologic and applied studies (see Particular Applications and Table 13.3).

Given the preponderance of older, sicker men in the VA system, the VA database is distinctly not representative of the US population or even of all US veterans, a majority of whom do not receive care within the VA system. However, the VA serves a high proportion of traditionally underrepresented and vulnerable groups, including the elderly and those with multiple co-morbidities, mental illness, disabilities, and lower socioeconomic status. The large and growing size of the VA population and retention of patients in the system, often until their death, facilitate large longitudinal studies within this special population of veterans.

Validity of Clinical Information

Epidemiologic studies in any of the EHR databases involve use of lists of codes, and sometimes algorithms, for specific medical conditions, drugs or other exposures of interest, and co-variates. Methods for deriving such code lists have been described [49,108]. The validity of such code lists and algorithms has been extensively studied in many of these databases. Studies of agreement between recording in the EHR and capture of data (e.g., prescription medications and specialist referrals) have been performed for certain databases [17,22,50,66,67,109]. Numerous studies have validated outcomes within EHR databases (see Table 13.3). For unvalidated outcomes of

Table 13.3 Examples of studies using EHR databases

Type of research	Setting or subject area	Sample of publications
Validation studies	BIFAP	Community-acquired pneumonia [189], ischemic stroke [190], meningioma [191], myocardial infarction [192], Stevens–Johnson syndrome and toxic epidermal necrolysis [193], and upper gastrointestinal bleeding [194]
	CPRD [17]	Atrial fibrillation [195], autism [196], cancer [197], cataract [198], chronic obstructive pulmonary disease [199–201], familial hypercholesterolemia [202], inflammatory bowel disease [203], liver injury [204], lymphoma [205], myocardial infarction and heart disease [206,207], pregnancy outcomes [68,69], pressure ulcers [208], psoriasis [48,209,210], psychosis [211], rheumatoid arthritis and juvenile idiopathic arthritis [212], Stevens–Johnson syndrome and toxic epidermal necrolysis [213], suicide [214], and venous thromboembolism [215]
	IQVIA DA databases [22]	Venous thromboembolism [216], general validation of pharmacoepidemiologic and pharmacoeconomic studies [22]
	LPD Italy	Heart failure, ischemic heart disease, hypertension, and type 2 diabetes [3]
	THIN [16]	Quality of cancer reporting [1,217], date of death and mortality reporting [53], hepatitis C virus infection [51], myocardial infarction [50], nonmelanomatous skin cancer [54], peptic ulcer disease [50,185], psoriasis [210], and stroke [50]
	VA	Acute kidney injury [218], fatty liver disease [219], heart failure [220], hepatocellular cancer [221], inflammatory bowel disease [222], myocardial infarction and related cardiac procedures [223], posttraumatic stress disorder [224], sepsis [225], and stroke [226]
Methodologic studies	Database-specific research (CPRD or THIN)	Timing and validity of diagnoses and outcomes relative to EHR-specific administrative dates [133,227,228] and clinical coding practices [229–231]; methods to impute missing drug information [134]; potential for misclassification and resultant bias due to missing data within free text [232,233], paper records, [234] and linked hospitalization records [235]
	Generalizable research using EHR databases (various)	Sources and types of bias [236–240]; novel study designs, such as self-controlled designs [241], regression discontinuity designs [242], and prevalent new-user cohort designs [243]; various analytic approaches for causal inference, such as propensity scores [244], high-dimensional propensity scores [245], simulation [246], marginal structural models [247,248], targeted maximum likelihood estimation [249], and instrumental variable analysis [250]; handling of missing data [251–253]; handling of repeated data [254,255]; and identification of outliers [256]

Table 13.3 (Continued)

Type of research	Setting or subject area	Sample of publications
Applied studies	General epidemiologic studies	Europe: representative incidence and prevalence studies [177,195,257–266] (e.g., shoulder complaints in primary care [267], newly diagnosed heart failure [268], bullous pemphigoid [269], and pemphigus vulgaris [269]); natural history of disease (e.g., irritable bowel syndrome [270]); risk of disease-related outcomes [260,271–279] (e.g., lymphoma among inflammatory bowel disease patients [205], myocardial infarction in patients with psoriasis [274], and complications of diabetes [280]); research on associated conditions [208,281] (e.g., obesity and liver disease [282]); patterns of diseases or symptoms [199]; rates of referral (e.g., chronic pelvic pain [283,284]); impact of geography [285,286] and pollution [287–289] on disease incidence and outcomes US VA: burden of illness associated with irritable bowel syndrome [290], military sexual trauma [291], patients awaiting major joint arthroplasty [292], and mental illness among veterans [293]
	Pharmacoepidemiologic studies	Europe: studies assessing risks [113,197,214,279,294–304] and outcomes [305–311] of medication (e.g., risk of myopathy and myalgias by statins [299]); safety and tolerability of medications [14,29,33,248,298,312–319]; studies of medication exposure and pregnancy outcomes [71,320,321]; reduction of morbidity or mortality by medication [306,322,323] and vaccinations [324]; persistence of medication use [325–327] (e.g., antihypertensives [328,329], bisphosphonates [330,331], and glaucoma therapies [332]); compliance and adherence [24,25,333,334]; physician's use of guidelines in prescribing medications [335–338] (e.g., antibiotics in children [339,340], antidepressants [341]); trends in prescribing [342–351]; device utilization [352,353], effectiveness [354], and safety [355,356] US VA: risks of myocardial infarction or musculoskeletal pain associated with bisphosphonates [357], dysglycemia with fluoroquinolones [358], neuropsychiatric adverse events with smoking cessation therapy [359], gastrointestinal bleeding with selective serotonin reuptake inhibitors [360], glucocorticoid-induced osteoporosis [361], and antipsychotic-associated mortality in dementia patients [362]
	Pharmacoeconomics, health services research, and pharmacovigilance	Europe: cost-effectiveness and safety of bisphosphonates [363], comparison of cost between glaucoma therapies [332,364], use and cost-effectiveness of long-term hormonal contraception [26,365], cost-effectiveness of treatment of gastroesophageal reflux disease [366]; health insurance-related barriers in access to and compliance with medicines [367–370]; healthcare utilization in fibromyalgia [371] and diabetes [30,372–376]; prescribing trends and their financial impact [377,378]; comparison of care of the elderly and nonelderly regarding symptoms concerning for ovarian cancer [379]; research on disparities and health outcomes [285,380–384]; variability in resource utilization and prescribing [385,386]; vaccination uptake and distribution [387–389]; impact of risk minimization measures [27,31] US VA: costs of erythropoietin therapy [390,391], treatment for metastatic prostate cancer [392], atrial fibrillation and stroke prevention [393], and healthcare costs of a collaborative intervention for chronic pain [394]; impact of clinical practice guidelines on quality of care [395] and adherence to the diabetes guidelines [396]

interest, one should strongly consider validating outcomes before or during a study to ensure that the specified diagnostic codes or algorithms reflect patients' true conditions. Codes or diagnoses that have not been validated may lead to spurious results and compromise a study's validity [50,51] (see also Chapter 37).

Accuracy of Drug Information

Electronic health record databases contain information on the name, strength, and quantity of prescribed drugs, which can be used to estimate their expected end date. In the UK, unlike in other countries, the prescription is the payment document. Although information is usually lacking on whether a prescribed drug has been dispensed from a UK pharmacy or taken by the patient, new prescriptions are generated and recorded when the current refills have been used and the patient requests a refill. A prespecified number of repeat prescriptions can be issued upon request of patients, after which a repeat clinical evaluation is required to ensure the prescribed therapy is still appropriate. The number of prespecified refills depends on factors such as patients' medical history and the drug in question.

Arianna, BIFAP, Pedianet, SIDIAP, and the VA databases contain information on outpatient drug dispensing in addition to the prescription data. VA databases also include detailed data on inpatient drug dispensing and administration. Of note, research using THIN has shown a high correspondence between issued and dispensed prescriptions except for a few selected drug classes; antipsychotics, drugs for malignancy, and immunosuppressants had lower redemption rates while anesthesia and vaccines were underreported as prescribed [110].

Ability to Access Original Health Records

Electronic health record databases contain information from patients' actual health records, which gives researchers insights into medical and social histories that are not possible with other types of databases (e.g., claims). For instance, researchers can access information about antecedent symptoms, prior medical conditions, family history, vital signs, physical exam finding, laboratory data, and prescribed medications, as detailed above. Particularly in closed medical systems where GPs and family pediatricians are the gatekeepers, health information in the EHR tends to be relatively comprehensive (see Europe and the United Kingdom, Overview of Healthcare Systems and Populations). Notably, some EHR databases permit access to anonymized free text data (e.g., IPCI), anonymized copies of paper records (e.g., THIN), or the entire EHR (e.g., VA), as well as access to clinicians or patients via surveys. These options allow researchers to verify information found elsewhere in a database or obtain additional supplemental data (see Data Collection and Structure). In published studies from the UK, response rates for health record requests have been greater than 80–90% [111–113].

Linkage to External Patient-Level Data

Many EHR databases may be linked to other health-related, patient-level information, thus extending the functionality and utility of EHR data. More than 75% of applications submitted to CPRD request the use of linked data (sometimes customized for a study) to augment the information available for research. The data source most commonly linked to CPRD and THIN is HES, which can provide data on hospital visits and stays, accident and emergency episodes, and tests by specialists, including imaging. The combination of data from primary care and HES facilitates research on conditions managed across multiple healthcare settings [114–117]. Linkage to official death records may improve the accuracy of mortality studies and validate data from general practice [118,119]. Researchers can link EHR data with other data sources, including disease (e.g.,

cancer) registries, mental health datasets, and socioeconomic and deprivation indices [120–124]. Furthermore, linkage of EHR data to individual patient-generated data is also possible, including patient-reported outcomes, environmental data, drug diaries, and biospecimens [125–127].

Electronic health record databases outside the UK also permit linkages to other data sources, including Arianna (e.g., claims data and registries on regional hospitalizations, national drug dispensing, and mortality), SIDIAP, and the VA (e.g., genetic information from the Million Veteran Program biobank [128] and administrative claims data from Medicare and Medicaid [92,129,130]). See Chapter 12 for information about other data sources that combine EHR and claims data.

Limitations

Incompleteness of Clinical Data

When using EHR databases for research, investigators rely on recording of patients' history and events by their clinicians and the health systems they work in. In clinical practice, human errors and omissions naturally occur, but systematic errors in recording can lead to bias. For example, in studies of the elderly, researchers can use available geriatric data to create a frailty score [131]. Notably, such data are likely to be selectively recorded for frailer patients [132]. Failure to account for such patterns of missingness could lead to bias. Although most EHR data are in electronic form, information received from outside sources (e.g., consultants or hospitals) in hard copy may not be fully captured if results are not manually reentered by clinicians. As with the above example, clinicians may be more likely to record laboratory or radiologic findings that are abnormal, but even abnormal results may not be documented reliably.

Because European EHR databases are designed to capture health information from primary care settings, they typically lack information from specialists. IQVIA DA databases record data from certain specialists, but these encounters are not linked directly to patients' primary care records. In other databases, recording of information from secondary care relies on communication between GPs and other clinicians. Researchers using THIN or CPRD may access more extensive and reliable data from other settings through linkage with HES data and other sources (see Linkage to External Patient-Level Data).

The nature of illness can affect the pattern of data recording. Unlike in administrative claims databases, codes for chronic diseases may be entered only once into EHR databases. For this reason, episodes of care involving acute events may be better recorded than chronic diseases [48,109,133]. Likewise, codes for chronic or inactive medical conditions may predate a study's follow-up or fixed baseline period. Failure to account for historical diagnoses could lead to misclassification of important study variables. Another consideration for studies using EHR data involves follow-up time and censoring: patients may transfer out of a given practice and entire practices may stop contributing patient-level data to the database. If such dropout relates both to exposure status and the outcome, bias could result.

While certain types of information can be found in EHR data and not administrative claims, some of these same variables often contain missing data. Examples include race and ethnicity, smoking and alcohol use, BMI, socioeconomic status, employment status, and occupation (see Tables 13.1 and 13.2 for database-specific details on percent recording). Unless incentives or other processes actively encourage recording (see Data Quality: Accuracy and Completeness), clinicians may more likely document these variables of interest (potential confounders) when they consider them relevant to patients' health. Of note, some databases, including CPRD, THIN, and SIDIAP, derive socioeconomic data from

patients' location of residence rather than their own finances.

Another source of missing data involves pediatric growth measurements in the UK. GPs routinely measure and record the height and weight of children in a paper record provided to families. However, GPs inconsistently document these same measurements electronically. As a result, longitudinal measures of pediatric growth are frequently incomplete in CPRD and THIN; as with other variables, recording of these measures may relate to other clinical characteristics. Recording of growth measurements in children is more comprehensive in Pedianet, although the overall size of Pedianet's pediatric population is smaller than that of either UK database.

Veterans in the VHA may receive healthcare outside the VHA either by choice (especially veterans aged 65 and older with Medicare coverage) or by necessity. For instance, veterans with urgent or emergent conditions are taken to the nearest hospital for care, as their VHA hospitals may be farther away or may lack a true emergency department. As a result, occurrences of acute conditions, such as myocardial infarction, stroke, and severe hypoglycemia, may be generally missed in inpatient data from VA hospitals, potentially resulting in missing outcome data. The frequent omission of acute outcomes of interest represents a major limitation of the VA database, making it less useful for some types of pharmacoepidemiologic studies when used as the only data source. Of note, in studies of veterans age 65 and older, one can overcome this limitation by linking VA data to Medicare claims data [91–93].

Incompleteness of Drug Data

Information on days' supply and daily dosage of prescribed medicines may not be explicitly recorded in EHR data. Methods are available for imputing days' supply [134]. Information obtained from the timing of repeat prescriptions or refills can inform the imputation of daily doses [135,136]. Additionally, algorithms have been developed to determine daily dosage and other drug data (e.g., frequency) from unstructured text [137–139]. Only a few databases (Arianna, BIFAP, Pedianet, and to some extent DA) specifically link prescribed drugs to a particular diagnosis. Without this information, one can refer to diagnoses recorded in or around encounters that correspond to prescribed drugs.

One must remember that prescribing records do not indicate which prescriptions were filled. Only a few of the EHR databases discussed (Arianna, BIFAP, Pedianet, SIDIAP, and the VA) also contain drug dispensing data, providing a more complete picture of drug utilization. Data on OTC drugs are frequently missing from EHR databases, but exceptions exist where healthcare systems pay for OTC drugs. For example, long-term use of certain OTC medications, such as aspirin and nonsteroidal antiinflammatory drugs, is recorded in UK databases [140] and Arianna [32,33]. The UK National Health Service provides free prescription items to certain segments of the population (e.g., patients over 60, children under 16, people in full-time education). Patients over age 60 also have free access to some chronically used nonprescription medications that are prescribed by GPs. These accommodations lead to more reliable usage data for OTC drugs in some UK populations. Additionally, Pedianet and SIDIAP capture prescribing and dispensing of drugs irrespective of coverage by the respective national health systems, leading to comprehensive recording of OTC medications. Notably, medication adherence is not well recorded in any setting; thus, documentation of prescribing or dispensing does not imply that patients actually take their medications (see Chapter 38 on Adherence).

In European EHR databases, data on medications restricted to specialist care,

dispensed from hospital pharmacies (e.g., biologics), and given during hospitalization or upon hospital discharge may be missing. In the UK, patients generally receive a limited quantity of medications upon hospital discharge, which in some cases may last two weeks. In the VA, certain medications are recorded in the EHR but are not accessible in the prescription databases, for example medications obtained from floor stock or administered acutely in acute care areas. The administration of other drugs may occasionally be incomplete due to work-arounds in documentation [141].

Complexity and Costs

The size and complexity of these databases require adequate computer hardware, software, and storage space as well as experienced data managers and analysts. These requirements carry costs in addition to the costs of using many databases themselves. Open-source software is available for managing and analyzing EHR data [142,143], including algorithms for use and analysis of free text [144,145].

Particular Applications

The aforementioned EHR databases have been a rich scientific setting for thousands of epidemiologic and pharmacoepidemiologic investigations. Research using these databases has included assessments of the incidence and natural history of disease; research on drug utilization, safety, and effectiveness; pharmacovigilance and signal detection; health economics research including cost-effectiveness; studies using case–control, cohort, longitudinal, self-controlled, and other designs; and a variety of methodologic studies. Examples of these studies are provided in Table 13.3. More comprehensive lists of publications using these databases may be found on their respective websites.

The Future

Electronic health records continue to evolve and expand in pursuit of delivering high-quality, high-value healthcare [146]. Interoperable EHR platforms and health information exchanges can enable broader access to and sharing of data across clinical settings. Such advances enable better coordination of care, reduce redundancy, and improve efficiency within and across healthcare systems. New technologies have improved communication not just among clinicians. Clinicians and their patients can now correspond electronically via the EHR and even interact in virtual clinical encounters through telemedicine. In some settings, patients can see their EHR data through patient-specific portals and upload their own data (e.g., patient-reported outcomes, consumer-wearable technology, "smart" digital technology) into the EHR. All these changes rely on and generate increasing volumes of data. Through advanced data analytics, healthcare systems can harness these data to learn from their patients, align and streamline processes of care, and improve patient outcomes – key objectives of learning health systems [147].

As EHR systems evolve, systems administrators, clinical informatics specialists, and end users must address important challenges. Missing data and variability in recording are common in EHRs. To optimize clinical care and maximize the potential for high-quality research, healthcare systems must implement approaches to ensure consistent and complete clinical documentation within the EHR. Future research will be needed to determine whether continuing education, financial incentive programs (e.g., QOF), targeted feedback, or other strategies can improve

patient outcomes while being cost-effective and sustainable.

In an age of ever-growing threats to data security, EHR databases require enhanced vigilance, technology, and standards to maintain individuals' privacy and confidentiality. Healthcare systems, industries, regulatory agencies, and governments must balance the potential societal benefits of broad access to and linking of health data with the personal risks to privacy and safety [148]. Contractual, practical, and technical barriers must be overcome to link and share data from disparate, proprietary EHR platforms.

The many advancements in EHR systems have important implications for the conduct of research. New technologic approaches, such as natural language processing, artificial intelligence, and other big-data analytics, have enabled researches to organize and use complex EHR data in novel ways [149–152]. Nonetheless, the clinical value of artificial intelligence-based methods to study medicines in observational settings remains unclear [153,154]. Large international research networks, such as ARITMO [29], SOS [155], and TEDDY [156] have demonstrated the capacity and power of international collaborations to use EHR databases for large-scale research on drug utilization and outcomes. Such collaborations lead to increased statistical power to study rare diseases, uncommonly used drugs, and rare outcomes; they also have greater external validity than research conducted within a single region or country. Linkage of EHRs with hospital data, administrative claims, and other data sources (e.g., patient registries) is increasingly important to maximize the advantages of each data source and minimize their respective limitations (see also Chapters 12 and 14) [92,116,157]. Furthermore, linkage of EHR data to patient-generated data and biospecimens adds to the possibilities for inquiry and discovery through population-representative, patient-centered research and molecular pharmacoepidemiology [128,158,159]. Researchers can also use the clinical networks and information infrastructure of EHR systems to conduct large pragmatic trials [160–162]. In addition, expansion of EHRs in low- and middle-income countries facilitates research and improvements in healthcare in areas of great need [163]. High-quality research conducted within EHR databases can favorably influence public policy and public health [164].

In the area of pharmacovigilance and regulation, EHR databases have an important role in postmarket evaluation. With rapid, nearly real-time analyses of recent population-based data, EHR databases are useful data sources to conduct postauthorization safety studies that monitor utilization and potential health risks of new drugs. Researchers, industries, and regulatory agencies may also use EHRs to quantify the impact of risk minimization measures, such as drug safety warnings issued by national drug agencies [165].

Through high-quality clinical and translational research and pharmacovigilance, EHR systems of the future will continue to serve as key platforms for answering important questions and improving the health of patients, communities, and populations.

Acknowledgment

The authors wish to thank the following individuals who provided information about specific databases: Arianna: Valentina Ientile; BIFAP: Miguel Gil García, Miguel Angel Maciá Martínez, Dolores Montero Corominas; HSD: Alessandro Pasqua, Iacopo Cricelli, Francesco Lapi; IPCI: Marcel de Wilde; Pedianet: Carlo Giaquinto, Luigi Cantarutti, Anna Cantarutti; SIDIAP: Bonaventura Bolíbar, Eduardo Hermosilla Pérez, Maria del Mar García Gil, Talita Duarte Salles; VA: Lucy Pandey, Kwan Hur. The authors also thank Molly E. Collins for her valuable editorial assistance.

References

1 Haynes K, Forde K, Schinnar R, Wong P, Strom B, Lewis J. Cancer incidence in The Health Improvement Network. *Pharmacoepidemiol Drug Saf* 2009; **18**: 730–6.

2 Peters SG, Buntrock JD. Big data and the electronic health record. *J Ambul Care Manage* 2014; **37**(3): 206–10.

3 Gini R, Schuemie MJ, Mazzaglia G, *et al*. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. *BMJ Open* 2016; **6**(12): e012413.

4 Rivero-Calle I, Pardo-Seco J, Aldaz P, *et al*. Incidence and risk factor prevalence of community-acquired pneumonia in adults in primary care in Spain (NEUMO-ES-RISK project). *BMC Infect Dis* 2016; **16**(1): 645.

5 van Ginneken E. Perennial health care reform – the long Dutch quest for cost control and quality improvement. *N Engl J Med* 2015; **373**(10): 885–9.

6 García-Armesto S, Abadía-Taira MB, Durán A, Hernández-Quevedo C, Bernal-Delgado E. Health systems in transition: Spain, health system review. //www.euro.who.int/_data/assets/pdf_file/0004/128830/e94549.pdf (accessed April 9, 2019).

7 Heath D. France's health care system. *CMAJ* 2008; **178**(5): 596.

8 Lopes C. Health care in France: facing hard choices. *CMAJ* 2007; **177**(10): 1167–9.

9 Rodwin V. The health care system under French national health insurance: lessons for health reform in the United States. *Am J Public Health* 2003; **93**(1): 31–7.

10 Schlette S, Lisac M, Blum K. Integrated primary care in Germany: the road ahead. *Int J Integr Care* 2009; **9**: e14.

11 Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997; **350**: 1097–9.

12 Lawson D, Sherman V, Hollowell J. The General Practice Research Database. Scientific and Ethical Advisory Group. *Q J Med* 1998; **91**: 445–52.

13 Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf* 2001; **10**: 379–83.

14 Wood L, Martinez C. The General Practice Research Database: role in pharmacovigilance. *Drug Saf* 2004; **27**(12): 871–81.

15 Wong I, Murray M. The potential of UK clinical databases in enhancing paediatric medication research. *Br J Clin Pharmacol* 2005; **59**(6): 750–5.

16 Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2007; **16**(4): 393–401.

17 Herrett E, Thomas S, Schoonen S, Smeeth L, Hall A. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69**(1): 4–14.

18 Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**(3): 827–36.

19 Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care* 2011; **19**(4): 251–5.

20 Carbonari DM, Saine ME, Newcomb CW, *et al*. Use of demographic and pharmacy data to identify patients included within both the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Pharmacoepidemiol Drug Saf* 2015; **24**(9): 999–1003.

21 Cai B, Xu W, Bortnichak E, Watson DJ. An algorithm to identify medical practices common to both the General Practice Research Database and The Health Improvement Network database. *Pharmacoepidemiol Drug Saf* 2012; **21**(7): 770–4.

22 Becher H, Kostev K, Schroder-Bernhardi D. Validity and representativeness of the Disease Analyzer patient database for use in pharmacoepidemiological and pharmacoeconomic studies. *Int J Clin Pract* 2009; **47**(10): 617–26.

23 Dietlein G, Schroder-Bernhardi D. Use of the Mediplus patient database in healthcare research. *Int J Clin Pharmacol Ther* 2002; **40**(3): 130–3.

24 Guglielmi V, Bellia A, Pecchioli S, *et al*. Effectiveness of adherence to lipid lowering therapy on LDL-cholesterol in patients with very high cardiovascular risk: a real-world evidence study in primary care. *Atherosclerosis* 2017; **263**: 36–41.

25 Vetrano DL, Bianchini E, Onder G, *et al*. Poor adherence to chronic obstructive pulmonary disease medications in primary care: role of age, disease burden and polypharmacy. *Geriatrics Gerontol Int* 2017; **17**(12): 2500–6.

26 Lapi F, Simonetti M, Cricelli I, Cricelli C, Cassano N, Vena GA. Prescription appropriateness of cyproterone acetate/ethinylestradiol in primary care: a population-based study in Italy. *Clin Drug Invest* 2017; **37**(8): 755–62.

27 Sultana J, Fontana A, Giorgianni F, *et al*. The effect of safety warnings on antipsychotic drug prescribing in elderly persons with dementia in the United Kingdom and Italy: a population-based study. *CNS Drugs* 2016; **30**(11): 1097–109.

28 Tocci G, Nati G, Cricelli C, *et al*. Prevalence and control of hypertension in the general practice in Italy: updated analysis of a large database. *J Human Hypertens* 2017; **31**(4): 258–62.

29 Trifiro G, de Ridder M, Sultana J, *et al*. Use of azithromycin and risk of ventricular arrhythmia. *CMAJ* 2017; **189**(15): E560–E568.

30 Trifiro G, Parrino F, Pizzimenti V, *et al*. The management of diabetes mellitus in patients with chronic kidney disease: a population-based study in southern Italy. *Clin Drug Invest* 2016; **36**(3): 203–12.

31 Viola E, Trifiro G, Ingrasciotta Y, *et al*. Adverse drug reactions associated with off-label use of ketorolac, with particular focus on elderly patients. An analysis of the Italian pharmacovigilance database and a population based study. *Expert Opin Drug Saf* 2016; **15**(Suppl 2): 61–7.

32 Ingrasciotta Y, Sultana J, Giorgianni F, *et al*. The burden of nephrotoxic drug prescriptions in patients with chronic kidney disease: a retrospective population-based study in Southern Italy. *PloS One* 2014; **9**(2): e89072.

33 Ingrasciotta Y, Sultana J, Giorgianni F, *et al*. Association of individual non-steroidal anti-inflammatory drugs and chronic kidney disease: a population-based case control study. *PloS On.* 2015; **10**(4): e0122899.

34 Sultana J, Musazzi UM, Ingrasciotta Y, *et al*. Medication is an additional source of phosphate intake in chronic kidney disease patients. *Nutr Metab Cardiovasc Dis* 2015; **25**(10): 959–67.

35 Guerriero F, Orlando V, Tari DU, *et al*. How healthy is community-dwelling elderly population? Results from Southern Italy. *Transl Med UniSa* 2015; **13**: 59–64.

36 Brauer R, Ruigomez A, Downey G, *et al*. Prevalence of antibiotic use: a comparison across various European health care data sources. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 11–20.

37 Ruigomez A, Brauer R, Rodriguez LA, *et al*. Ascertainment of acute liver injury in two European primary care databases. *Eur J Clin Pharmacol* 2014; **70**(10): 1227–35.

38 Jordan K, Clarke AM, Symmons DP, *et al*. Measuring disease prevalence: a comparison

of musculoskeletal disease using four general practice consultation databases. *Br J Gen Pract* 2007; **57**(534): 7–14.

39 Nightingale AL, Farmer RDT, de Vries CS. Systemic lupus erythematosus prevalence in the UK: methodological issues when using the General Practice Research Database to estimate frequency of chronic relapsing-remitting disease. *Pharmacoepidemiol Drug Saf* 2007; **16**(2): 144–51.

40 Maguire A, Blak B, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* 2009; **18**: 76–83.

41 Martin-Merino E, Huerta-Alvarez C, Prieto-Alhambra D, Montero-Corominas D. Cessation rate of anti-osteoporosis treatments and risk factors in Spanish primary care settings: a population-based cohort analysis. *Arch Osteoporos* 2017; **12**(1): 39.

42 Cao X, Ejzykowicz F, Ramey DR, *et al*. Impact of switching from high-efficacy lipid-lowering therapies to generic simvastatin on LDL-C levels and LDL-C goal attainment among high-risk primary and secondary prevention populations in the United Kingdom. *Clin Therapeut* 2015; **37**(4): 804–15.

43 Kostev K, Parhofer KG, Dippel FW. Prevalence of high-risk cardiovascular patients with therapy-resistant hypercholesterolemia. *Cardiovasc Endocrinol* 2017; **6**(2): 81–5.

44 Ramos R, Ballo E, Marrugat J, *et al*. Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol* 2012; **65**(1): 29–37.

45 Vinagre I, Mata-Cases M, Hermosilla E, *et al*. Control of glycemia and cardiovascular risk factors in patients with type 2 diabetes in primary care in Catalonia (Spain). *Diabetes Care* 2012; **35**(4): 774–9.

46 Salvador-Gonzalez B, Mestre-Ferrer J, Soler-Vila M, *et al*. Chronic kidney disease in hypertensive subjects >/=60 years treated in primary care. *Nefrologia* 2017; **37**(4): 406–14.

47 Gunathilake W, Song S, Sridharan S, Fernando DJ, Idris I. Cardiovascular and metabolic risk profiles in young and old patients with type 2 diabetes. *QJM* 2010; **103**(11): 881–4.

48 Gelfand J, Margolis D, Dattani H. The UK General Practice Research Database. In: Strom B, ed. Pharmacoepidemiology, 4th edn. Chichester: John Wiley & Sons, 2005, pp. 337–46.

49 Dave S, Peterson I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009; **18**: 704–7.

50 Lewis J, Schinnar R, Bilker W, Wang X, Strom B. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiology Drug Saf* 2007; **16**: 393–401.

51 Lo Re V, Haynes K, Forde K, Localio A, Schinnar R, Lewis J. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. *Pharmacoepidemiol Drug Saf* 2009; **18**: 807–14.

52 Khan N, Perera R, Harper S, Rose P. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract* 2010; **5**(11): 1.

53 Hall G. Validation of death and suicide recording on the THIN UK primary care database. *Pharmacoepidemiol Drug Saf* 2009; **18**: 120–31.

54 Meal A, Leonardi-Bee J, Smith C, Hubbard R, Bath-Hextall F. Validation of THIN data for non-melanoma skin cancer. *Qual Primary Care* 2008; **16**: 49–52.

55 Khan N, Harrison S, Rose P. Validity of diagnostic coding within the General Practice Research Database: a systemic review. *Br J Gen Pract* 2010; **60**(572): e128–e136.

56 Scott FI, Mamtani R, Haynes K, Goldberg DS, Mahmoud NN, Lewis JD. Validation of a coding algorithm for intra-abdominal surgeries and adhesion-related complications in an electronic medical records database. *Pharmacoepidemiol Drug Saf* 2016; **25**(4): 405–12.

57 deLusignan S, Stephens P, Naeema N, Majeed A. Does feedback improve the quality of computerized medical records in primary care? *J Am Med Informatics Assoc* 2002; **9**(4): 395–403.

58 Doran T, Fullwood C, Gravelle H, *et al*. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006; **355**: 375–84.

59 Crawley D, Ng A, Mainous A, Majeed A, Millett C. Impact of pay for performance on quality of chronic disease management by social class group in England. *J R Soc Med* 2009; **102**: 103–7.

60 McDonald R, Roland M. Pay for performance in primary care in England and California: comparison of unintended consequences. *Ann Fam Med* 2009; **7**: 121–7.

61 Kontopantelis E, Springate D, Reeves D, Ashcroft DM, Valderas JM, Doran T. Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework. *BMJ* 2014; **348**: g330.

62 Campbell S, Reeves D, Kontopantelis E, Middleton E, Siddald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med* 2007; **357**(2): 181–90.

63 Forbes LJ, Marchand C, Doran T, Peckham S. The role of the Quality and Outcomes Framework in the care of long-term conditions: a systematic review. *Br J Gen Pract* 2017; **67**(664): e775–e784.

64 Marshall M, Roland M. The future of the Quality and Outcomes Framework in England. *BMJ* 2017; **359**: j4681.

65 Health and Social Care Information Centre. National Dementia and Antipsychotic Prescribing Audit: National Summary Report. https://digital.nhs.uk/catalogue/PUB06624 (accessed April 9, 2019).

66 Jick SS, Kaye JA, Vasilakis-Scaramozza C, *et al*. Validity of the general practice research database. *Pharmacotherapy* 2003; **23**(5): 686–9.

67 Jick H, Jick S, Derby L. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991; **302**(6779): 766–8.

68 Charlton R, Cunnington M, de Vries C, Weil J. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf* 2008; **31**(1): 39–51.

69 Devine S, West S, Andrews E, Tennis P, Hammad T, Eaton S. The identification of pregnancies within the general research database. *Pharmacoepidemiol Drug Saf* 2010; **19**(1): 45–50.

70 Martín-Merino E, Ruigómez A, Johansson S, Wallander M, García-Rodriguez L. Study of a cohort of patients newly diagnosed with depression in general practice: prevalence, incidence, comorbidity, and treatment patterns. *Prim Care Compan J Clin Psychiatry* 2010; **12**(1).

71 Tata L, Lewis S, Mckeever T, *et al*. Congenital malformation in children born to women with asthma. *Pharmacoepidemiol Drug Saf* 2007; **16**(S271).

72 Hardy J, Holford T, Hall G, Bracken M. Strategies for identifying pregnancies in the automated medical records of General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2004; **13**: 749–59.

73 Hardy J, Leaderer B, Holford T, Hall G, Bracken M. Safety of medications prescribed before and during early pregnancy in a cohort of 81,975 mothers from the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2006; **15**(8): 555–64.

74 McKeever T, Lewis S, Smith C, Collins J, Heatlie H, Frischer M. Siblings, multiple births, and the incidence of allergic disease: a birth cohort study using the West Midlands general practice research database. *Thorax* 2001; **56**(10): 758–62.

75 Joseph RM, Movahedi M, Dixon WG, Symmons DP. Smoking-related mortality in patients with early rheumatoid arthritis: a retrospective cohort study using the Clinical Practice Research Datalink. *Arthritis Care Res (Hoboken)* 2016; **68**(11): 1598–606.

76 Lane DA, Skjoth F, Lip GYH, Larsen TB, Kotecha D. Temporal trends in incidence, prevalence, and mortality of atrial fibrillation in primary care. *J Am Heart Assoc* 2017; **6**(5): ii.

77 Stewart D, Han L, Doran T, McCambridge J. Alcohol consumption and all-cause mortality: an analysis of general practice database records for patients with long-term conditions. *J Epidemiol Community Healt.* 2017; **71**(8): 729–35.

78 Scott FI, Horton DB, Mamtani R, *et al*. Administration of antibiotics to children before age 2 years increases risk for childhood obesity. *Gastroenterology* 2016; **151**(1): 120–9 e125.

79 Ogdie A, Maliha S, Shin D, *et al*. Cause-specific mortality in patients with psoriatic arthritis and rheumatoid arthritis. *Rheumatology (Oxford)* 2017; **56**(6): 907–11.

80 Lester H. The UK Quality and Outcomes Framework. *BMJ* 2008; **337**: a2095.

81 Sutton M, Elder R, Guthrie B, Watt G. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Economics* 2010; **19**(1): 1–13.

82 Wise J. Framework for NHS annual "quality accounts" is launched. *BMJ* 2009; **339**: b3890.

83 Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol Drug Saf* 2013; **22**(12): 1357–61.

84 Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open* 2013; **3**(9): e003389.

85 Marston L, Carpenter JR, Walters KR, *et al*. Smoker, ex-smoker or non-smoker? The validity of routinely recorded smoking status in UK primary care: a cross-sectional study. *BMJ Open* 2014; **4**(4): e004958.

86 Lewis J, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf* 2004; **13**: 437–41.

87 US Department of Veterans Affairs. VA History. www.va.gov/about_va/vahistory.asp (accessed April 9, 2019)

88 Panangala SV. *Veterans Medical Care: FY 2011 Appropriations*. CRS Report for Congress. Washington, DC: DIANE Publishing, 2010.

89 National Center for Veterans Analysis and Statistics. NCVAS Pocket Cards. www.va.gov/vetdata/pocketcard/index.asp (accessed April 9, 2019).

90 Thomas KS, Cote D, Makineni R, *et al*. Change in VA community living centers 2004–2011: shifting long-term care to the community. *J Aging Soc Policy* 2018; **30**(2): 93–108.

91 Gellad WF. The Veterans Choice Act and dual health system use. *J Gen Intern Med* 2016; **31**(2): 153–4.

92 Thorpe JM, Thorpe CT, Gellad WF, *et al*. Dual health care system use and high-risk prescribing in patients with dementia: a national cohort study. *Ann Intern Med* 2017; **166**(3): 157–63.

93 Gellad WF, Thorpe JM, Zhao X, *et al*. Impact of dual use of Department of Veterans Affairs and Medicare Part D drug benefits on potentially unsafe opioid use. *Am J Public Health* 2018; **108**(2): 248–55.

94 US Department of Veteran Affairs. National Patient Care Database (NPCD). catalog.data.gov/dataset/national–patient–care–database–npcd (access via log-in only).

95 Robb MA, Racoosin JA, Worrall C, Chapman S, Coster T, Cunningham FE. Active surveillance of postmarket medical product safety in the Federal Partners' Collaboration. *Med Care* 2012; **50**(11): 948–53.

96 US Department of Veterans Affairs. 172VA10P2: VHA Corporate Data Warehouse–VA (79 FR 4377). 2014.

97 US Department of Veterans Affairs. Corporate Data Warehouse (CDW). www.data.va.gov/dataset/corporate–data–warehouse–cdw (accessed April 9, 2019).

98 Tarlov E. Ascertaining veterans' vital status and dates of death: the VHA Vital Status File and other data sources for mortality ascertainment in VA. www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=857 (accessed April 9, 2019)

99 Aspinall SL, Sales MM, Good CB, *et al*. Pharmacy benefits management in the Veterans Health Administration revisited: a decade of advancements, 2004–2014. *J Manage Care Specialty Pharmacy* 2016; **22**(9): 1058–63.

100 Smith MW, Joseph GJ. Pharmacy data in the VA health care system. *Med Care Res Rev* 2003; **60**(3 Suppl): 92S–123S.

101 US Department of Veterans Affairs. *Bar Code Medication Administration (BCMA): Manager's User Manual*. Washington, DC: US Department of Veterans Affairs, 2004.

102 Patterson OV, Freiberg MS, Skanderson M, Fodeh SJ, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. *BMC Cardiovasc Disord* 2017; **17**(1): 151.

103 Schroeck FR, Patterson OV, Alba PR, *et al*. Development of a natural language processing engine to generate bladder cancer pathology data for health services research. *Urology* 2017; **110**: 84–91.

104 Kazis LE. The Veterans SF-36 Health Status Questionnaire: development and application in the Veterans Health Administration. *Medical Outcomes Trust Monitor* 2000; **5**(1): 1–14.

105 Gebregziabher M, Zhao Y, Axon N, Gilbert G, Echols C, Egede L. Lessons learned in dealing with missing race data: an empirical investigation. *J Biomet Biostat* 2012; **3**(3): 1–5.

106 Szklo M. Population-based cohort studies. *Epidemiol Rev* 1998; **20**(1).

107 Rothman K, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott-Raven, 1988, pp. 93–115.

108 Dregan A, Grieve A, van Staa T, Gulliford MC, eCRT Research Team. Potential application of item–response theory to interpretation of medical codes in electronic patient records. *BMC Med Res Methodol* 2011; **11**: 168.

109 Hollowell J. The General Practice Research Database: quality of morbidity data. *Popul Trends* 1997; **87**: 36–40.

110 Prescribing and Primary Care Services of the NHS Information Centre. Prescribing Compliance: A Review of the Proportion of Prescriptions Dispensed. https://digital.nhs.uk/data-and-information/publications/statistical/prescribing-compliance/prescribing-compliance-a-review-of-the-proportion-of-prescriptions-dispensed (accessed April 10, 2019).

111 Arellano F, Arana A, Wentworth C, Fernandez-Vidaurre C, Schlienger R, Conde E. Lymphoma among patients with atopic dermatitis and/or treated with topical immunosuppressants in the United Kingdom. *J Allergy Clin Immunol* 2009; **123**(5): 1111–16.

112 Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality evaluated database of primary care data. *Informat Primary Care* 2004; **12**(3): 171–7.

113 Van Staa T, Abenhaim L, Cooper C, Zhang B, Leufkens H. The use of a large pharmacoepidemiological database to study exposure to oral corticosteroids and risk of fractures: validation of study population and results. *Pharmacoepidemiol Drug Saf* 2000; **9**(5): 359–66.

114 Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. *BMC Health Serv Res* 2012; **12**: 392.

115 Herrett E, Shah AD, Boggon R, *et al*. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ.* 2013; **346**: f2350.

116 Millett ER, Quint JK, de Stavola BL, Smeeth L, Thomas SL. Improved incidence estimates from linked vs. stand-alone electronic health records. *J Clin Epidemiol* 2016; **75**: 66–9.

117 Berni E, de Voogd H, Halcox JP, *et al*. Risk of cardiovascular events, arrhythmia and all–cause mortality associated with clarithromycin versus alternative antibiotics prescribed for respiratory tract infections: a retrospective cohort study. *BMJ Open* 2017; **7**(1): e013398.

118 Gallagher AM, Williams T, Leufkens HG, de Vries F. The impact of the choice of data source in record linkage studies estimating mortality in venous thromboembolism. *PLoS One* 2016; **11**(2): e0148349.

119 Morgan C, Webb RT, Carr MJ, *et al*. Incidence, clinical management, and mortality risk following self harm among children and adolescents: cohort study in primary care. *BMJ* 2017; **359**: j4351.

120 Iyen-Omofoman B, Hubbard RB, Smith CJ, *et al*. The distribution of lung cancer across sectors of society in the United Kingdom: a study using national primary care data. *BMC Public Health* 2011; **11**: 857.

121 Curtis EM, van der Velde R, Moon RJ, *et al*. Epidemiology of fractures in the United Kingdom 1988–2012: variation with age, sex, geography, ethnicity and socioeconomic status. *Bone* 2016; **87**: 19–26.

122 Bailey SE, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: a prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract* 2017; **67**(659): e405–e413.

123 Morris M, Woods LM, Bhaskaran K, Rachet B. Do pre-diagnosis primary care consultation patterns explain deprivation-specific differences in net survival among women with breast cancer? An examination of individually-linked data from the UK West Midlands cancer registry, national screening programme and Clinical Practice Research Datalink. *BMC Cancer* 2017; **17**(1): 155.

124 Young GJ, Harrison S, Turner EL, *et al*. Prostate-specific antigen (PSA) testing of men in UK general practice: a 10-year longitudinal cohort study. *BMJ Open* 2017; **7**(10): e017729.

125 Moore E, Chatzidiakou L, Jones RL, *et al*. Linking e-health records, patient-reported symptoms and environmental exposure data to characterise and model COPD exacerbations: protocol for the COPE study. *BMJ Open* 2016; **6**(7): e011330.

126 Joseph RM, Soames J, Wright M, Sultana K, van Staa TP, Dixon WG. Supplementing electronic health records through sample collection and patient diaries: a study set within a primary care research database. *Pharmacoepidemiol Drug Saf* 2018; **27**(2): 239–42.

127 O'Meara H, Carr DF, Evely J, *et al*. Electronic health records for biological sample collection: feasibility study of statin-induced myopathy using the Clinical Practice Research Datalink. *Br J Clin Pharmacol* 2014; **77**(5): 831–8.

128 Gaziano JM, Concato J, Brophy M, *et al*. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016; **70**: 214–23.

129 Abraham NS, Castillo DL, Hartman C. National mortality following upper gastrointestinal or cardiovascular events in older veterans with recent nonsteroidal anti–inflammatory drug use. *Aliment Pharmacol Ther* 2008; **28**(1): 97–106.

130 Wong ES, Rinne ST, Hebert PL, Cook MA, Liu CF. Hospital distance and readmissions among VA–Medicare dual-enrolled veterans. *J Rural Health* 2016; **32**(4): 377–86.

131 Clegg A, Bates C, Young J, *et al*. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* 2018; **47**(2): 319.

132 Sultana J, Fontana A, Giorgianni F, *et al*. Can information on functional and cognitive status improve short-term mortality risk prediction among community-dwelling older people? A cohort study using a UK primary care database. *Clin Epidemiol* 2018; **10**: 31–9.

133 Lewis J, Bilker W, Weinstein R, Strom B. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2005; **14**: 443–51.

134 Lum KJ, Newcomb CW, Roy JA, *et al*. Evaluation of methods to estimate missing days' supply within pharmacy data of the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Eur J Clin Pharmacol* 2017; **73**(1): 115–23.

135 Meid AD, Heider D, Adler JB, *et al*. Comparative evaluation of methods approximating drug prescription durations in claims data: modeling, simulation, and application to real data. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1434–42.

136 Stovring H, Pottegard A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1451–9.

137 Shah AD, Martinez C. An algorithm to derive a numerical daily dose from unstructured text dosage instructions. *Pharmacoepidemiol Drug Saf* 2006; **15**(3): 161–6.

138 Karystianis G, Sheppard T, Dixon WG, Nenadic G. Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database. *BMC Med Inform Decis Mak* 2016; **16**: 18.

139 Lu CC, Leng J, Cannon GW, *et al*. The use of natural language processing on narrative medication schedules to compute average weekly dose. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1414–24.

140 Yang YX, Hennessy S, Propert K, Hwang W, Sedarat A, Lewis J. Chronic proton pump inhibitor therapy and the risk of colorectal cancer. *Gastroenterology* 2007; **133**: 748–54.

141 Staggers N, Iribarren S, Guo JW, Weir C. Evaluation of a BCMA's electronic medication administration record. *Western J Nurs Res* 2015; **37**(7): 899–921.

142 Egbring M, Kullak-Ublick GA, Russmann S. Phynx: an open source software solution supporting data management and web-based patient-level data review for drug safety studies in the general practice research database and other health care databases. *Pharmacoepidemiol Drug Saf* 2010; **19**(1): 38–44.

143 Springate DA, Parisi R, Olier I, Reeves D, Kontopantelis E. rEHR: an R package for manipulating and analysing Electronic Health Record data. *PloS One* 2017; **12**(2): e0171784.

144 Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting

diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PloS One* 2012; **7**(1): e30412.

145 Shah AD, Martinez C, Hemingway H. The freetext matching algorithm: a computer program to extract diagnoses and causes of death from unstructured text in electronic health records. *BMC Med Inform Decis Mak* 2012; **12**: 88.

146 Evans RS. Electronic health records: then, now, and in the future. *Yearb Med Inform* 2016; **Suppl 1**: S48–61.

147 Friedman C, Rubin J, Brown J, *et al*. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc* 2015; **22**(1): 43–50.

148 Rumbold JM, Pierscionek B. The effect of the General Data Protection Regulation on medical research. *J Med Internet Res* 2017; **19**(2): e47.

149 Nelson SD, Lu CC, Teng CC, *et al*. The use of natural language processing of infusion notes to identify outpatient infusions. *Pharmacoepidemiol Drug Saf* 2015; **24**(1): 86–92.

150 Walker AM, Zhou X, Ananthakrishnan AN, *et al*. Computer-assisted expert case definition in electronic health records. *Int J Med Inform* 2016; **86**: 62–70.

151 Garvin JH, Kalsy M, Brandt C, *et al*. An evolving ecosystem for natural language processing in Department of Veterans Affairs. *J Med Syst* 2017; **41**(2): 32.

152 Trivedi G, Pham P, Chapman WW, Hwa R, Wiebe J, Hochheiser H. NLPReViz: an interactive tool for natural language processing on clinical text. *J Am Med Inform Assoc* 2018; **25**(1): 81–7.

153 Trifiro G, Sultana J, Bate A. From Big Data to smart data for pharmacovigilance: the role of healthcare databases and other emerging sources. *Drug Saf* 2018; **41**(2): 143–9.

154 Bate A, Reynolds RF, Caubel P. The hope, hype and reality of Big Data for pharmacovigilance. *Ther Adv Drug Saf* 2018; **9**(1): 5–11.

155 Arfe A, Scotti L, Varas-Lorenzo C, *et al*. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case–control study. *BMJ* 2016; **354**: i4857.

156 Hsia Y, Neubert A, Sturkenboom MC, *et al*. Comparison of antiepileptic drug prescribing in children in three European countries. *Epilepsia* 2010; **51**(5): 789–96.

157 Chu TP, Shah A, Walker D, Coleman MP. Where are the opportunities for an earlier diagnosis of primary intracranial tumours in children and young adults? *Eur J Paediatr Neurol* 2017; **21**(2): 388–95.

158 Wing K, Douglas I, Bhaskaran K, *et al*. Development of predictive genetic tests for improving the safety of new medicines: the utilization of routinely collected electronic health records. *Drug Discov Today* 2014; **19**(4): 361–6.

159 Reade S, Spencer K, Sergeant JC, *et al*. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR Mhealth Uhealth* 2017; **5**(3): e37.

160 Van Staa TP, Dyson L, McCann G, *et al*. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014; **18**(43): 1–146.

161 Juszczyk D, Charlton J, McDermott L, *et al*. Electronically delivered, multicomponent intervention to reduce unnecessary antibiotic prescribing for respiratory infections in primary care: a cluster randomised trial using electronic health records – REDUCE Trial study original protocol. *BMJ Open* 2016; **6**(8): e010892.

162 Lederle FA, Cushman WC, Ferguson RE, Brophy MT, Fiore Md LD. Chlorthalidone versus hydrochlorothiazide: a new kind of Veterans Affairs cooperative study. *Ann Intern Med* 2016; **165**(9): 663–4.

163 Aminpour F, Sadoughi F, Ahamdi M. Utilization of open source electronic health record around the world: a systematic review. *J Res Med Sci* 2014; **19**(1): 57–64.

164 Oyinlola JO, Campbell J, Kousoulis AA. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Serv Res* 2016; **16**: 299.

165 Goedecke T, Morales DR, Pacurariu A, Kurz X. Measuring the impact of medicines regulatory interventions – systematic review and methodological considerations. *Br J Clin Pharmacol* 2018; **84**(3): 419–33.

166 Kaguelidou F, de Bie S, Verhamme K, *et al*. New quality indicators for paediatric antibiotic prescribing in primary care: a population based cohort study in the United Kingdom, Italy and the Netherlands from 1995–2010. *Arch Dis Child* 2016; **101**: e1–e1.

167 Mathur R, Bhaskaran K, Chaturvedi N, *et al*. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health* 2014; **36**(4): 684–92.

168 Guglielmi V, Bellia A, Bianchini E, *et al*. Drug interactions in users of tablet vs. oral liquid levothyroxine formulations: a real-world evidence study in primary care. *Endocrine* 2018; **59**(3): 585–92.

169 Gadroen K, Kemmeren JM, Bruijning-Verhagen PC, *et al*. Baseline incidence of intussusception in early childhood before rotavirus vaccine introduction, the Netherlands, January 2008 to December 2012. *Euro Surveillance* 2017; **22**(25).

170 Zamora A, Masana L, Comas-Cufi M, *et al*. Familial hypercholesterolemia in a European Mediterranean population – prevalence and clinical data from 2.5 million primary care patients. *J Clin Lipid* 2017; **11**(4): 1013–22.

171 US Department of Veteran Affairs. Pharmacy Benefits Management Services. www.pbm.va.gov/nationalformulary.asp (accessed April 10, 2019).

172 Sturkenboom MC, Dieleman JP, Picelli G, *et al*. Prevalence and treatment of hypertensive patients with multiple concomitant cardiovascular risk factors in The Netherlands and Italy. *J Human Hypertens* 2008; **22**(10): 704–13.

173 Garcia-Gil M, Hermosilla E, Prieto-Alhambra D, *et al*. Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). *Inform Prim Care* 2011; **19**(3): 135–45.

174 Valkhoff VE, Schade R, te Jong GW, *et al*. Population-based analysis of non-steroidal anti-inflammatory drug use among children in four European countries in the SOS project: what size of data platforms and which study designs do we need to assess safety issues? *BMC Pediatr* 2013; **13**: 192.

175 Rodriguez LA, Tolosa LB, Ruigomez A, Johansson S, Wallander MA. Rheumatoid arthritis in UK primary care: incidence and prior morbidity. *Scand J Rheumatol* 2009; **38**(3): 173–7.

176 Mulnier H, Seaman H, Raleigh V, Soedamah-Muthu S, Colhoun H, Lawrenson R. Mortality in people with type 2 daibetes in the UK. *Diabet Med* 2006; **23**: 516–21.

177 Neimann A, Shin D, Wang X, Margolis D, Troxel A, Gelfand J. Prevalence of cardiovascular risk factors in patients with psoriasis. *J Am Acad Dermatol* 2006; **55**: 829–35.

178 Delaney J, Daskalopoulou S, Brophy J, Steele R, Opatrny L, Suissa S. Lifestyle variables and the risk of myocardial infarction in the General Practice Research Database. *BMC Cardiovasc Disord* 2007; **7**(38): 1–8.

179 Cazzola M, Calzetta L, Lauro D, *et al*. Asthma and COPD in an Italian adult population: role of BMI considering the

smoking habit. *Respir Med* 2013; **107**(9): 1417–22.

180 Lapi F, Cassano N, Pegoraro V, *et al*. Epidemiology of chronic spontaneous urticaria: results from a nationwide, population-based study in Italy. *Br J Dermatol* 2016; **174**(5): 996–1004.

181 Afonso AS, Verhamme KM, Sturkenboom MC, Brusselle GG. COPD in the general population: prevalence, incidence and survival. *Respir Med* 2011; **105**(12): 1872–84.

182 Farmer R, Lawrenson R, Todd J, Williams T, MacRae K. Oral contraceptives and venous thromboembolic disease. Analysis of the UK General Practice Research Database and the UK MediPlus Database. *Human Reprod Update* 1999; **5**(6): 688–706.

183 Szathowski L, Lewis S, McNeill A, Coleman T. Is smoking status routinely recorded when patients register with a new GP? *Fam Pract* 2010; **27**(6): 673–5

184 Langley T, Szatkowski L, Gibson J, *et al*. Validation of the Health Improvement Network (THIN) primary care database for monitoring prescriptions for smoking cessation medications. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 586–90.

185 Cai S, Garcia Rodriguez L, Masso-Gonzalez E, Hernandez-Diaz S. Uncomplicated peptic ulcer in the UK: trends from 1997–2005. *Aliment Pharmacol Ther* 2009; **30**(10): 1039–48.

186 Baena-Diez JM, Garcia-Gil M, Comas-Cufi M, *et al*. Association between chronic immune-mediated inflammatory diseases and cardiovascular risk. *Heart* 2018;**104**(2): 119–26.

187 Cheung K, Aarts N, Noordam R, *et al*. Antidepressant use and the risk of suicide: a population-based cohort study. *J Affect Disord.* 2015; **174**: 479–84.

188 Gonzalez-Perez A, Garcia-Rodriguez L. Prostate cancer risk among men with diabetes mellitus. *Cancer Causes Control* 2005; **16**: 1055–8.

189 Saiz LC, Garjon J, Gorricho J, Erviti J, Gil-Garcia MJ, Martin-Merino E. Validation and incidence of community-acquired pneumonia in patients with type 2 diabetes in the BIFAP database. *Epidemiol Infect* 2017; **145**(14): 3056–64.

190 Garcia-Poza P, de Abajo FJ, Gil MJ, Chacon A, Bryant V, Garcia-Rodriguez LA. Risk of ischemic stroke associated with non-steroidal anti-inflammatory drugs and paracetamol: a population-based case–control study. *J Thromb Haemost* 2015; **13**(5): 708–18.

191 Gil M, Oliva B, Timoner J, Macia MA, Bryant V, de Abajo FJ. Risk of meningioma among users of high doses of cyproterone acetate as compared with the general population: evidence from a population-based cohort study. *Br J Clin Pharmacol* 2011; **72**(6): 965–8.

192 de Abajo FJ, Gil MJ, Garcia Poza P, *et al*. Risk of nonfatal acute myocardial infarction associated with non-steroidal antiinflammatory drugs, non-narcotic analgesics and other drugs used in osteoarthritis: a nested case–control study. *Pharmacoepidemiol Drug Saf* 2014; **23**(11): 1128–38.

193 Martin-Merino E, de Abajo FJ, Gil M. Risk of toxic epidermal necrolysis and Stevens–Johnson syndrome associated with benzodiazepines: a population-based cohort study. *Eur J Clin Pharmacol* 2015; **71**(6): 759–66.

194 de Abajo FJ, Gil MJ, Bryant V, Timoner J, Oliva B, Garcia-Rodriguez LA. Upper gastrointestinal bleeding associated with NSAIDs, other drugs and interactions: a nested case–control study in a new general practice database. *Eur J Clin Pharmacol* 2013; **69**(3): 691–701.

195 Ruigomez A, Johannson S, Wallander M, Rodriguez L. Incidence of chronic atrial fibrillation in general practice and its treatment pattern. *J Clin Epidemiol* 2002; **55**: 358–63.

196 Fombonne E, Heavey L, Smeeth L, Rodrigues L, Cook C, Smith P. Validation of the diagnosis of autism in general practitioner records. *BMC Public Health* 2004; **4**(5).

197 Jick H. Calcium-channel blockers and risk of cancer. *Lancet* 1997; **349**(9066): 1699–700.

198 Derby L, Maier W. Risk of cataract among users of intranasal corticosteroids. *J Allergy Clin Immunol* 2000; **105**(5): 912–16.

199 Soriano J, Maier W, Visick G, Pride N. Validation of general practitioner-diagnosed COPD in the UK General Practice Research Database. *Eur J Epidemiol* 2001; **17**: 1075–80.

200 Rothnie KJ, Mullerova H, Hurst JR, *et al*. Validation of the recording of acute exacerbations of COPD in UK primary care electronic healthcare records. *PLoS One* 2016; **11**(3): e0151357.

201 Rothnie KJ, Chandan JS, Goss HG, Mullerova H, Quint JK. Validity and interpretation of spirometric recordings to diagnose COPD in UK primary care. *Int J Chron Obstruct Pulmon Dis* 2017; **12**: 1663–8.

202 Weng SF, Kai J, Andrew Neil H, Humphries SE, Qureshi N. Improving identification of familial hypercholesterolaemia in primary care: derivation and validation of the familial hypercholesterolaemia case ascertainment tool (FAMCAT). *Atherosclerosis* 2015; **238**(2): 336–43.

203 Lewis J, Brensinger C, Bilker W, Strom B. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf* 2002; **11**: 211–18.

204 Wing K, Bhaskaran K, Smeeth L, *et al*. Optimising case detection within UK electronic health records: use of multiple linked databases for detecting liver injury. *BMJ Open* 2016; **6**(9): e012102.

205 Lewis J, Brensinger C, Bilker W, Deren J, Vaughn D, Strom B. Inflammatory bowel disease is not associated with an increased risk of lymphoma. *Gastroenterology* 2001; **121**: 1080–7.

206 Hammad TMM, Feight A, Iyasu S, Dal Pan G. Determining the predictive value of Read/OXMIS codes to identify incident acute myocardial infarction in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008; **17**(12): 1197–201.

207 Reeves D, Springate DA, Ashcroft DM, *et al*. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. *BMJ Open* 2014; **4**(4): e004952.

208 Margolis D, Bilker W, Knauss J, Baumgarten M, Strom B. The incidence and prevalence of pressure ulcers among elderly patients in general medical practice. *Ann Epidemiol* 2002; **12**(5): 321–5.

209 Gelfand J, Weinstein R, Porter S, Neimann A, Berlin J, Margolis D. Prevalence and treatment of psoriasis in the United Kingdom: a population-based study. *Arch Dermatol* 2005; **141**(12): 1537–41.

210 Seminara N, Abuabara K, Shin D, *et al*. Prevalence, treatment, and severity of psoriasis in The Health Improvement Network (THIN) – a population based study. *International Conference on Pharmacoepidemiology Conference,* 2010 (Abstract).

211 Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis of psychosis on general practice computer system. *BMJ* 1993; **307**(6895): 32–4.

212 Thomas S, Edwards C, Smeeth L, Cooper C, Hall A. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? *Arthritis Rheum* 2008; **59**(9): 1314–21.

213 Frey N, Bircher A, Bodmer M, Jick SS, Meier CR, Spoendlin J. Validation of Stevens–

Johnson syndrome or toxic epidermal necrolysis diagnoses in the Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 429–36.

214 Jick S, Dean A, Jick H. Antidepressants and suicide. *BMJ* 1995; **310**(6974): 215–18.

215 Lawrenson R, Todd J, Leydon G, Williams T, Farmer R. Validation of the diagnosis of venous thromboembolism in general practice database studies. *Pharmacoepidemiol Drug Saf* 2000; **49**(6): 591–6.

216 Todd J, Lawrenson R, Farmer R, Williams T, Leydon G. Venous thromboembolic disease and combined oral contraceptives: a re-analysis of the MediPlus database. *Human Reprod* 1999; **14**(6): 1500–5.

217 Arellano F, Conde E, Wenworth C, Schlienger R, Fernandez-Vidaurre C, Arana A. Validation of cases of lymphoma in THIN. *Pharmacoepi Drug Saf* 2008; **17**: S87–8.

218 Lafrance JP, Miller DR. Acute kidney injury associates with increased long-term mortality. *J Am Soc Nephrol* 2010; **21**(2): 345–52.

219 Husain N, Blais P, Kramer J, *et al*. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. *Aliment Pharmacol Ther* 2014; **40**(8): 949–54.

220 Mahajan SM, Burman P, Newton A, Heidenreich PA. A Validated risk model for 30-day readmission for heart failure. *Stud Health Technol Inform* 2017; **245**: 506–10.

221 Sada Y, Hou J, Richardson P, El-Serag H, Davila J. Validation of case finding algorithms for hepatocellular cancer from administrative data and electronic health records using natural language processing. *Med Care* 2016; **54**(2): e9–14.

222 Thirumurthi S, Chowdhury R, Richardson P, Abraham NS. Validation of ICD-9-CM diagnostic codes for inflammatory bowel disease among veterans. *Digest Dis Sci* 2010; **55**(9): 2592–8.

223 Petersen LA, Wright S, Normand SL, Daley J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. *J Gen Intern Med* 1999; **14**(9): 555–8.

224 Gravely AA, Cutting A, Nugent S, Grill J, Carlson K, Spoont M. Validity of PTSD diagnoses in VA administrative data: comparison of VA administrative PTSD diagnoses to self-reported PTSD Checklist scores. *J Rehab Res Dev* 2011; **48**(1): 21–30.

225 Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's Affairs hospital discharge databases coded serious bacterial infections accurately. *J Clin Epidemiol* 2007; **60**(4): 397–409.

226 Niesner K, Murff HJ, Griffin MR, et al. Validation of VA administrative data algorithms for identifying cardiovascular disease hospitalization. *Epidemiology* 2013; **24**(2): 334–5.

227 Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(1): 64–9.

228 Sammon CJ, Petersen I. Backdating of events in electronic primary health care data: should one censor at the date of last data collection. *Pharmacoepidemiol Drug Saf* 2016; **25**(4): 378–84.

229 Leite A, Andrews NJ, Thomas SL. Assessing recording delays in general practice records to inform near real-time vaccine safety surveillance using the Clinical Practice Research Datalink (CPRD). *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 437–45.

230 Tate AR, Dungey S, Glew S, Beloff N, Williams R, Williams T. Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study

using the CPRD database. *BMJ Open* 2017; **7**(1): e012905.

231 Tate AR, Martin AG, Murray-Thomas T, Anderson SR, Cassell JA. Determining the date of diagnosis – is it a simple matter? The impact of different approaches to dating diagnosis on estimates of delayed care for ovarian cancer in UK primary care. *BMC Med Res Methodo.* 2009; **9**: 42.

232 Ford E, Nicholson A, Koeling R, *et al*. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol* 2013; **13**: 105.

233 Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case–control study. *BMJ Open* 2016; **6**(5): e011664.

234 Charlton RA, Weil JG, Cunnington MC, de Vries CS. Identifying major congenital malformations in the UK General Practice Research Database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf* 2010; **33**(9): 741–50.

235 Rothnie KJ, Mullerova H, Thomas SL, *et al*. Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. *Clin Epidemiol* 2016; **8**: 771–82.

236 MacDonald TM, Morant SV, Goldstein JL, Burke TA, Pettitt D. Channelling bias and the incidence of gastrointestinal haemorrhage in users of meloxicam, coxibs, and older, non–specific non-steroidal anti-inflammatory drugs. *Gut* 2003; **52**(9): 1265–70.

237 Card TR, Solaymani-Dodaran M, Hubbard R, Logan RF, West J. Is an internal comparison better than using national data when estimating mortality in longitudinal studies? *J Epidemiol Community Health* 2006; **60**(9): 819–21.

238 Matok I, Azoulay L, Yin H, Suissa S. Immortal time bias in observational studies of drug effects in pregnancy. *Birth Defects Res A Clin Mol Teratol* 2014; **100**(9): 658–62.

239 Renoux C, dell'Aniello S, Brenner B, Suissa S. Bias from depletion of susceptibles: the example of hormone replacement therapy and the risk of venous thromboembolism. *Pharmacoepidemiol Drug Saf* 2017; **26**(5): 554–60.

240 Ankarfeldt MZ, Thorsted BL, Groenwold RH, Adalsteinsson E, Ali MS, Klungel OH. Assessment of channeling bias among initiators of glucose-lowering drugs: a UK cohort study. *Clin Epidemiol* 2017; **9**: 19–30.

241 Wang S, Linkletter C, Maclure M, *et al*. Future cases as present controls to adjust for exposure trend bias in case-only studies. *Epidemiology* 2011; **22**(4): 568–74.

242 Geneletti S, O'Keeffe AG, Sharples LD, Richardson S, Baio G. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Stat Med* 2015; **34**(15): 2334–52.

243 Suissa S, Moodie EE, dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 459–68.

244 Morant SV, Pettitt D, MacDonald TM, Burke TA, Goldstein JL. Application of a propensity score to adjust for channelling bias with NSAIDs. *Pharmacoepidemiol Drug Saf* 2004; **13**(6): 345–53.

245 Toh S, Garcia Rodriguez LA, Hernan MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf* 2011; **20**(8): 849–57.

246 Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *J Clin Epidemiol* 2007; **60**(7): 686–95.

247 Delaney JA, Daskalopoulou SS, Suissa S. Traditional versus marginal structural models to estimate the effectiveness of beta-blocker use on mortality after myocardial infarction. *Pharmacoepidemiol Drug Saf* 2009; **18**(1): 1–6.

248 Ali MS, Groenwold RH, Belitser SV, *et al*. Methodological comparison of marginal structural model, time-varying Cox regression, and propensity score methods: the example of antidepressant use and the risk of hip fracture. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 114–21.

249 Pang M, Schuster T, Filion KB, Eberg M, Platt RW. Targeted maximum likelihood estimation for pharmacoepidemiologic research. *Epidemiology* 2016; **27**(4): 570–7.

250 Boef AG, Souverein PC, Vandenbroucke JP, *et al*. Instrumental variable analysis as a complementary analysis in studies of adverse effects: venous thromboembolism and second-generation versus third-generation oral contraceptives. *Pharmacoepidemiol Drug Saf* 2016; **25**(3): 317–24.

251 Delaney JA, Moodie EE, Suissa S. Validating the effects of drug treatment on blood pressure in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008; **17**(6): 535–45.

252 Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 618–26.

253 Welch CA, Petersen I, Bartlett JW, *et al*. Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Stat Med* 2014; **33**(21): 3725–37.

254 Crowther MJ, Lambert PC, Abrams KR. Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach. *BMC Med Res Methodol* 2013; **13**: 146.

255 Sheppard T, Tamblyn R, Abrahamowicz M, Lunt M, Sperrin M, Dixon WG. A comparison of methods for estimating the temporal change in a continuous variable: example of HbA1c in patients with diabetes. *Pharmacoepidemiol Drug Saf* 2017; **26**(12): 1474–82.

256 Welch C, Petersen I, Walters K, *et al*. Two-stage method to remove population- and individual-level outliers from longitudinal data in a primary care database. *Pharmacoepidemiol Drug Saf* 2012; **21**(7): 725–32.

257 Linsell L, Dawson J, Zondervan K, *et al*. Prospective study of elderly people comparing treatments following first primary care consultation for a symptomatic hip or knee. *Fam Pract* 2005; **22**(1): 118–25.

258 Bath-Hextall F, Leonardi-Bee J, Smith C, Meal A, Hubbard R. Trends in incidence of skin basal cell carcinoma: additional evidence from a UK primary care database study. *Int J Cancer* 2007; **121**(9): 2105–8.

259 Van Staa T, Selby P, Leufkens H, Lyles K, Sprafka J, Cooper C. Incidence and natural history of Paget's disease of bone in England and Wales. *J Bone Min Res* 2002; **17**(3): 465–71.

260 Fischer L, Schlienger R, Matter C, Jick H, Meier C. Effect of rheumatoid arthritis or systemic lupus erythematosus on the risk of first-time acute myocardial infarction. *Am J Cardiol* 2004; **93**(2): 198–200.

261 Gupta D, Hansell A, Nichols T, Duong T, Ayres J, Strachan D. Epidemiology of pneumothorax in England. *Thorax* 2000; **55**(8): 666–71.

262 Humes D, Solaymani-Dodaran M, Fleming K, Simpson J, Spiller R, West J. A population-based study of perforated diverticular disease incidence and associated mortality. *Gastroenterology* 2009; **136**(4): 1198–205.

263 Punekar Y, Sheikh A. Establishing the incidence and prevalence of clinician-diagnosed allergic conditions in children and

adolescents using routinely collected data from general practices. *Clin Exp Allergy* 2009; **39**(8): 1209–16.

264 Souverein P, Webb D, Petri H, Weil J, Staa TV, Egberts T. Incidence of fractures among epilepsy patients: a population-based retrospective cohort study in the General Practice Research Database. *Epilepsia* 2005; **46**(2): 304–10.

265 Somers E, Thomas S, Smeeth L, Schoonen W, Hall A. Incidence of systemic lupus erythematosus in the United Kingdom; 1990–1999. *Arthritis Rheum* 2007; **57**(4): 612–18.

266 Davies A, Smeeth L, Grundy E. Contribution of changes in incidence and mortality to trends in the prevalence of coronary heart disease in the UK: 1996–2005. *Eur Heart J* 2007; **28**: 2142–7.

267 Linsell L, Dawson J, Zondervan K, *et al*. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patters of diagnosis and referral. *Rheumatology* 2006; **45**: 215–21.

268 Johansson S, Wallander MA, Ruigomez A, Garcia Rodriguez LA. Incidence of newly diagnosed heart failure in UK general practice. *Eur J Heart Fail* 2001; **3**(2): 225–31.

269 Langan S, Smeeth L, Hubbard R, Fleming K, Smith C, West J. Bullous pemphigoid and pemphigus vulgaris – incidence and mortality in the UK: population based cohort study. *BMJ* 2008; **337**(a180).

270 Ruigomez A, Wallander MA, Johansson S, Garcia Rodriguez LA. One-year follow-up of newly diagnosed irritable bowel syndrome patients. *Aliment Pharmacol Ther* 1999; **13**(8): 1097–102.

271 Kaplan GG, Hubbard J, Korzenik J, *et al*. The inflammatory bowel diseases and ambient air pollution: a novel association. *Am J Gastroenterol* 2010; **105**(11): 2412–19.

272 Solaymani-Dodaran M, Aithal G, Card T, West J. Risk of cardiovascular and cerebrovascular events in primary biliary cirrhosis: a population-based cohort. *Am J Gastroenterol* 2008; **103**(11): 2784–8.

273 Clayton T, Thompson M, Meade T. Recent respiratory infection and risk of cardiovascular disease: case–control study through a general practice database. *Eur Heart J* 2008; **29**: 96–103.

274 Gelfand J, Neimann A, Shin D, Wang X, Margolis D, Troxel A. Risk of myocardial infarction in patients with psoriasis. *JAMA* 2006; **296**(14): 1735–42.

275 Pujades-Rodriguez M, Smith C, Hubbard R. Inhaled corticosteroids and the risk of fracture in chronic obstructive pulmonary disease. *Q J Med* 2007; **100**(8): 509–17.

276 Breart G, Cooper C, Meyer O, Speirs C, Deltour N, Reginster J. Osteoporosis and venous thromboembolism: a retrospective cohort study in the UK General Practice Research Database. *Osteoporos Int* 2010; **21**(7): 1181–7.

277 Hernán M, Jick S, Logroscino G, Olek M, Ascherio A, Jick H. Cigarette smoking and the progression of multiple sclerosis. *Brain* 2005; **128**(6): 1461–5.

278 Lewis N, Logan R, Hubbard R, West J. No increase in risk of fracture, malignancy or mortality in dermatitis herpetiformis: a cohort study. *Aliment Pharmacol Ther* 2008; **27**(11): 1140–7.

279 Van Staa T, Geusens P, Pols H, Laet CD, Leufkens H, Cooper C. A simple score for estimating the long-term risk of fracture in patients using oral glucocorticoids. *Q J Med* 2005; **98**: 191–8.

280 Huerta C, Zhao SZ, Garcia Rodriguez LA. Risk of acute liver injury in patients with diabetes. *Pharmacotherapy* 2002; **22**(9): 1091–6.

281 Davey G, Sedgwick P, Maier W, Visick G, Strachan D, Anderson H. Association between migraine and asthma: matched case–control study. *Br J Gen Pract* 2002; **52**(482): 723–7.

282 Meier C, Krahenbuhl S, Schlienger R, Jick H. Association between body mass index and liver disorders: an epidemiological study. *J Hepatol* 2002; **37**: 741–7.

283 Zondervan K, Yudkin P, Vessey M, Dawes M, Barlow D, Kennedy S. Patterns of diagnosis and referral in women consulting for chronic pelvic pain in UK primary care *Br J Obstet Gynaecol* 1999; **106**(11): 1156–61.

284 Zondervan K, Yudkin P, Vessey M, Dawes M, Barlow D, Kennedy S. Prevalence and incidence of chronic pelvic pain in primary care: evidence from a national general practice database. *Br J Obstet Gynaecol* 1999; **106**(11): 1149–55.

285 Perry DC, Bruce CE, Pope D, Dangerfield P, Platt MJ, Hall AJ. Legg–Calvé–Perthes disease in the UK: geographic and temporal trends in incidence reflecting differences in degree of deprivation in childhood. *Arthritis Rheum* 2012; **64**(5): 1673–9.

286 Shah A, Prieto-Alhambra D, Hawley S, *et al*. Geographic variation in secondary fracture prevention after a hip fracture during 1999–2013: a UK study. *Osteoporos Int* 2017; **28**(1): 169–78.

287 Hajat S, Haines A, Goubet SA, Atkinson RW, Anderson HR. Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. *Thorax* 1999; **54**(7): 597–605.

288 Atkinson RW, Carey IM, Kent AJ, van Staa TP, Anderson HR, Cook DG. Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort. *Occup Environ Med* 2015; **72**(1): 42–8.

289 Carey IM, Anderson HR, Atkinson RW, *et al*. Traffic pollution and the incidence of cardiorespiratory outcomes in an adult cohort in London. *Occup Environ Med* 2016; **73**(12): 849–56.

290 Spiegel BM. Burden of illness in irritable bowel syndrome: looking beyond the patient. *Clin Gastroenterol Hepatol* 2013; **11**(2): 156–7.

291 Haskell SG, Gordon KS, Mattocks K, *et al*. Gender differences in rates of depression, PTSD, pain, obesity, and military sexual trauma among Connecticut War Veterans of Iraq and Afghanistan. *J Womens Health* 2010; **19**(2): 267–71.

292 Kelly KD, Voaklander D, Kramer G, Johnston DW, Redfern L, Suarez-Almazor ME. The impact of health status on waiting time for major joint arthroplasty. *J Arthroplasty* 2000; **15**(7): 877–83.

293 Maynard C, Batten A, Liu CF, Nelson K, Fihn SD. The burden of mental illness among veterans: use of VHA health care services by those with service-connected conditions. *Med Care* 2017; **55**(11): 965–9.

294 Blumentals W, Brown R, Gomez-Caminero A. Antihypertensive treatment and erectile dysfunction in a cohort of type II diabetes patients. *Int J Impot Res* 2003; **15**(5): 314–17.

295 Kornegay C, Vasilakis-Scaramozza C, Jick H. Incident diabetes associated with antipsychotic use in the United Kingdom general practice research database. *J Clin Psychiatry* 2002; **63**(9): 758–62.

296 Van Staa T, Abenhaim L, Monette J. Rates of hypoglycemia in users of sulfonylureas. *J Clin Epidemiol* 1997; **50**(6): 735–41.

297 Langman M, Eichler E, Mavros P, Watson D, Kong S. Initiation of antihypertensive therapy among new users of cyclooxygenase-2-selective and nonselective NSAIDs. *Int J Clin Pharmacol Ther* 2004; **42**(5): 260–6.

298 Mochenhaupt M, Messenheimer J, Tennis P, Schlingmann J. Risk of Stevens–Johnson syndrome and toxic epidermal necrolysis in new users of antiepileptics. *Neurology* 2005; **64**: 1134–8.

299 Molokhia M, McKeigue P, Curcin V, Majeed A. Statin induced myopathy and myalgia: time trend analysis and comparison of risk associated with statin class from 1991–2006. *PLoS One* 2008; **3**(6): e2522.

300 Zhang H, Garcia-Rodriguez L, Hernandez-Diaz S. Antibiotic use and the risk of lung cancer. *Cancer Epidemiol* 2008; **17**(6): 1308–15.

301 Becker C, Jick S, Meier C. Use of statins and the risk of Parkinson's disease: a retrospective case–control study in the UK. *Drug Saf* 2008; **31**(5): 399–407.

302 Castellsague J, García-Rodríguez L, Duque A, Pérez S. Risk of serious skin disorders among users of oral antifungals: a population-based study. *BMC Dermatol* 2002; **2**: 14.

303 Meier C, Derby L, Jick S, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 1999; **281**(5): 427–31.

304 Meier C, Wilcock K, Jick S. The risk of severe depression, psychosis or panic attacks with prophylactic antimalarials. *Drug Saf* 2004; **27**(3): 203–13.

305 Edwards C, Cooper C, Fisher D, Field M, Van Staa T, Arden N. The importance of the disease process and disease-modifying antirheumatic drug treatment in the development of septic arthritis in patients with rheumatoid arthritis. *Arthritis Rheum* 2007; **57**(7): 1151–7.

306 Bulugahapitiya U, Siyambalapitiya S, Sithole J, Fernando D, Idris I. Age threshold for vascular prophylaxis by aspirin in patients without diabetes. *Heart* 2008; **94**: 1429–32.

307 Gore M, Sadosky A, Leslie D, Sheehan A. Medication for treating neuropathic pain in patients with diabetes: a study using the UK and Germany Mediplus databases. *Pain Pract* 2008; **8**(4): 253–62.

308 Simons W, Vinod H, Gerber R, Bolinder B. Does rapid transition to insulin therapy in subjects with newly diagnosed type 2 diabetes mellitus benefit glycaemic control and diabetes-related complications? A German population-based study. *Exper Clin Endocrinol Diabetes* 2006; **114**(9): 520–6.

309 Smeeth L, Douglas I, Hall A, Hubbard R, Evans S. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. *Br J Clin Pharmacol* 2008; **67**(1): 99–109.

310 Thomas M, Williams A. Are outcomes the same with all dry powder inhalers? *Int J Clin Pract* 2005; **59**(Suppl 149): 33–5.

311 Zhang Q, Thomas M, Wisniewski T, Kocevar V, Price D. Treatment and outcomes in patients with asthma and allergic rhinitis in the UK. *Allergy Immunol.* 2007; **142**: 318–28.

312 McCarthy S, Cranswick N, Potts L, Taylor E, Wong I. Mortality associated with attention-deficit hyperactivity disorder (ADHD) drug treatment: a retrospective cohort study of children, adolescents and young adults using the general practice research database. *Drug Saf* 2009; **32**(11): 1089–96.

313 Hubbard R, Lewis S, West J, *et al*. Bupropion and the risk of sudden death: a self-controlled case-series analysis using The Health Improvement Network. *Thorax* 2005; **60**(10): 848–50.

314 Gaus W, Westendorf J, Diebow R, Kieser K. Identification of adverse drug reactions by evaluation of a prescription database, demonstrated for "risk of bleeding". *Methods Informat Med* 2005; **44**(5): 697–703.

315 Langman M, Kong S, Zhang Q, Kahler K, Finch E. Safety and patient tolerance of standard and slow-release formulations of NSAIDs. *Pharmacoepidemiol Drug Saf* 2003; **12**: 61–6.

316 Margolis D, Hoffstad O, Strom B. Association between serious ischemic cardiac outcomes and medications used to treat diabetes. *Pharmacoepidemiol Drug Saf* 2008; **17**: 753–9.

317 Tata L, Fortune P, Hubbard R, *et al*. Does concurrent prescription of selective serotonin reuptake inhibitors and non-steroidal anti-inflammatory drugs substantially increase the risk of upper

gastrointestinal bleeding? *Aliment Pharmacol Ther* 2005; **22**: 175–81.

318 de Bie S, Coloma PM, Ferrajolo C, *et al*. The role of electronic healthcare record databases in paediatric drug safety surveillance: a retrospective cohort study. *Br J Clin Pharmacol* 2015; **80**(2): 304–14.

319 Giner-Soriano M, Roso-Llorach A, Vedia Urgell C, *et al*. Effectiveness and safety of drugs used for stroke prevention in a cohort of non-valvular atrial fibrillation patients from a primary care electronic database. *Pharmacoepidemiol Drug Saf* 2017; **26**(1): 97–107.

320 Wood S, Jick H, Sauve R. The risk of stillbirth in pregnancies before and after the onset of diabetes. *Diabet Med* 2003; **20**: 703–7.

321 Zomerdijk IM, Ruiter R, Houweling LM, *et al*. Isotretinoin exposure during pregnancy: a population-based study in The Netherlands. *BMJ Open* 2014; **4**(11): e005602.

322 Garcia Rodriguez LA, Varas C, Patrono C. Differential effects of aspirin and non-aspirin nonsteroidal antiinflammatory drugs in the primary prevention of myocardial infarction in postmenopausal women. *Epidemiology* 2000; **11**(4): 382–7.

323 Gonzalez ELM, Rodriguez LAG. Proton pump inhibitors reduce the long-term risk of recurrent upper gastrointestinal bleeding: an observational study. *Aliment Pharmacol Therapeut* 2008; **28**(5): 629–37.

324 Schembri S, Winter J, MacDonald T. Influenza and pneumococcal vaccination protect against all cause mortality in patients with chronic obstructive pulmonary diseases. ISPE Abstract, Copenhagen, 2008.

325 Arellano FM, May C, Verma A, Cid J. Patterns of discontinuation, dose modification and switching among users of prescription non-steroidal anti-inflammatory drugs (NSAID) including cyclo-oxygenase 2 inhibitors (COX–2) in the UK population. *Pharmacoepidemiol Drug Saf* 2005; **14**(S1): 3.

326 Gopal M, Haynes K, Bellamy S, Arya L. Discontinuation rates of anticholinergic medications used for the treatment of lower urinary tract symptoms. *Obstet Gynecol* 2008; **112**(6): 1311–18.

327 Lyu R, Govoni M, Ding Q, *et al*. Treatment persistence among patients with rheumatoid disease (RA, AS, PsA) treated with subcutaneous biologics in Germany. *Rheumatol Int* 2016; **36**(1): 143–53.

328 Hasford J, Schroder-Bernhardi D, Rottenkolber M, Kostev K, Dietlein G. Persistence with antihypertensive treatments: results of a 3-year follow up cohort study. *Eur J Clin Pharmacol* 2007; **63**: 1055–61.

329 Suarez A, Staffa J, Fletcher P, Jones J. Reason for discontinuation of newly prescribed antihypertensive medications: methods of a pilot study using computerized patient records. *Pharmacoepidemiol Drug Saf* 2000; **9**: 405–16.

330 Cramer J, Gold D, Silverman S, Lewiecki E. A systematic review of persistence and compliance with bisphosphonates for osteoporosis. *Osteoporos Int* 2007; **18**(8): 1023–31.

331 Van Staa T, Geusens P, Leufkens H, Cooper C. Persistence to bisphosphonate treatment in actual clinical practice. *Bone* 2005; **36**(S2): S412.

332 Lafuma A, Berdeaux G. Costs and persistence of carbonic anhydrase inhibitor versus alpha-2 agonists, associated with beta-blockers, in glaucoma and ocular hypertension: an analysis of the UK GPRD database. *Curr Med Res Opin* 2008; **24**(5): 1519–27.

333 Schroder-Bernhardi D, Dietlein G. Compliance with prescription recommendations by physicians in practices. *Int J Clin Pharmacol Therapeut J* 2001; **39**(11): 477–9.

334 Ferrajolo C, Arcoraci V, Sullo MG, *et al*. Pattern of statin use in southern italian

primary care: can prescription databases be used for monitoring long-term adherence to the treatment? *PloS One* 2014; **9**(7): e102146.

335 Mikuls T, Farrar J, Bilker W, Fernandes S, Saag K. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). *Rheumatology* 2005; **44**(8): 1038–42.

336 Thompson PL, Gilbert RE, Long PF, Saxena S, Sharland M, Wong IC. Has UK guidance affected general practitioner antibiotic prescribing for otitis media in children? *J Public Health* 2008; **30**(4): 479–86.

337 Judge A, Wallace G, Prieto-Alhambra D, Arden NK, Edwards CJ. Can the publication of guidelines change the management of early rheumatoid arthritis? An interrupted time series analysis from the United Kingdom. *Rheumatology* 2015; **54**(12): 2244–8.

338 Kendrick T, Stuart B, Newell C, Geraghty AW, Moore M. Did NICE guidelines and the Quality Outcomes Framework change GP antidepressant prescribing in England? Observational study with time trend analyses 2003–2013. *J Affect Disord* 2015; **186**: 171–7.

339 Ferrajolo C, Verhamme KM, Trifiro G, *et al*. Antibiotic-induced liver injury in paediatric outpatients: a case–control study in primary care databases. *Drug Saf* 2017; **40**(4): 305–15.

340 Thompson P, Spyridis N, Sharland M, *et al*. Changes in clinical indications for community antibiotic prescribing for children in the UK from 1996 to 2006: will the new NICE prescribing guidance on upper respiratory tract infections just be ignored? *Arch Dis Child* 2009; **94**: 337–40.

341 Murray M, Thompson M, Santosh P, Wong I. Effects of the committee on safety of medicines advice on antidepressant prescribing to children and adolescents in the UK. *Drug Saf* 2005; **28**(12): 1151–7.

342 Munson J, Kreider M, Chen Z, Christie J, Kimmel S. Factors associated with the use of corticosteroids in the initial management of idiopathic pulmonary fibrosis. *Pharmacoepidemiol Drug Saf* 2010; **19**(7): 756–62.

343 Arellano F, Yood M, Wentworth C, *et al*. Use of cyclo-oxygenase 2 inhibitors (COX-2) and prescription non-steroidal anti-inflammatory drugs (NSAIDS) in UK and USA populations. Implications for COX-2 cardiovascular profile. *Pharmacoepidemiol Drug Saf* 2006; **15**: 861–72.

344 Boyc K, Yurgin N. Trends in the prescription of antidiabetic medications in France: evidence from primary care physicians. *Adv Ther* 2007; **24**(4): 803–14.

345 Sittle R, Nuijten M, Nautrup B. Patterns of dosage changes with transdermal buprenorphine and transdermal fentanyl for the treatment of noncancer and cancer pain: a retrospective data analysis in Germany. *Clin Therapeut* 2006; **28**(8): 1144–54.

346 Dietlein G, Schroder-Bernhardi D. Doctors' prescription behaviour regarding dosage recommendations for preparations of kava extracts. *Pharmacoepidemiol Drug Saf* 2003; **12**(5): 417–21.

347 Schroder-Bernhardi D, Dietlein G. Lipid-lowering therapy: do hospitals influence the prescribing behavior of general practitioners? *Int J Clin Pharmacol Therapeut* 2002; **40**(7): 317–21.

348 MacDonald T, Morant S. Prevalence and treatment of isolated and concurrent hypertension and hypercholesterolaemia in the United Kingdom. *Br J Clin Pharmacol* 2008; **65**(5): 775–86.

349 Turner S, Thomas M, Ziegenweidt Jv, Price D. Prescribing trends in asthma: a longitudinal observational study. *Arch Dis Child* 2009; **94**(1): 16–22.

350 Edwards C, Arden N, Fisher D, Saperia J, Reading I, Van Staa T. The changing use of disease-modifying anti-rheumatic drugs in

individuals with rheumatoid arthritis from the United Kingdom General Practice Research Database. *Rheumatology* 2005; **44**(11): 1394–8.

351 Suh D, Hunsche E, Shin H, Mavros P. Co-prescribing of proton pump inhibitors among chronic users of NSAIDs in the UK. *Rheumatology* 2008; **47**: 458–63.

352 Culliford DJ, Maskell J, Kiran A, *et al*. The lifetime risk of total hip and knee arthroplasty: results from the UK general practice research database. *Osteoarthritis Cartilage* 2012; **20**(6): 519–24.

353 Cea Soriano L, Wallander MA, Andersson S, Filonenko A, Garcia Rodriguez LA. Use of long-acting reversible contraceptives in the UK from 2004 to 2010: analysis using The Health Improvement Network Database. *Eur J Contracept Reprod Health Care* 2014; **19**(6): 439–47.

354 Guest JF, Fuller GW, Vowden P. Clinical outcomes and cost-effectiveness of three different compression systems in newly-diagnosed venous leg ulcers in the UK. *J Wound Care* 2017; **26**(5): 244–54.

355 Lalmohamed A, MacGregor AJ, de Vries F, Leufkens HG, van Staa TP. Patterns of risk of cancer in patients with metal-on-metal hip replacements versus other bearing surface types: a record linkage study between a prospective joint registry and general practice electronic health records in England. *PloS One* 2013; **8**(7): e65891.

356 Parker SE, Jick SS, Werler MM. Intrauterine device use and the risk of pre-eclampsia: a case–control study. *Br J Obstet Gynaecol* 2016; **123**(5): 788–95.

357 Caplan L, Pittman CB, Zeringue AL, *et al*. An observational study of musculoskeletal pain among patients receiving bisphosphonate therapy. *Mayo Clin Proc* 2010; **85**(4): 341–8.

358 Aspinall SL, Good CB, Jiang R, McCarren M, Dong D, Cunningham FE. Severe dysglycemia with the fluoroquinolones: a class effect? *Clin Infect Dis* 2009; **49**(3): 402–8.

359 Cunningham FE, Hur K, Dong D, *et al*. A comparison of neuropsychiatric adverse events during early treatment with varenicline or a nicotine patch. *Addiction* 2016; **111**(7): 1283–92.

360 Ayers K, Waljee AK, Saini SD. Short-term use of selective serotonin reuptake inhibitors increases risk of gastrointestinal bleeding in men with psychiatric diagnoses. *Gastroenterology* 2014; **147**(5): 1173–4.

361 Caplan L, Hines AE, Williams E, *et al*. An observational study of glucocorticoid-induced osteoporosis prophylaxis in a national cohort of male veterans with rheumatoid arthritis. *Osteoporos Int* 2011; **22**(1): 305–15.

362 Kales HC, Valenstein M, Kim HM, *et al*. Mortality risk in patients with dementia treated with antipsychotics versus other psychiatric medications. *Am J Psychiatry* 2007; **164**(10): 1568–76; quiz 1623.

363 Van Staa T, Geusens P, Leufkens H, Cooper C. An individualised model of cost-effectiveness of bisphosphonates in elderly women. *Bone* 2005; **36**(S2): S138.

364 Lafuma A, Berdeaux G. Costs and effectiveness of travoprost versus a dorzolamide + timolol fixed combination in first-line treatment of glaucoma: analysis conducted on the United Kingdom General Practitioner Research Database. *Curr Med Res Opin* 2007; **23**(12): 3009–16.

365 Varney S, Guest J. Relative cost effectiveness of Depo-Provera, Implanon, and Mirena in reversible long-term hormonal contraception in the UK. *Pharmacoeconomics* 2004; **22**(17): 1141–51.

366 Plumb J, Edwards S. Cost effectiveness analysis of proton pump inhibitors compared to omeprazole in the healing of reflux oesophagitis. *J Med Economics* 2002; **5**: 25–38.

367 Krobot K, Kaufman J, Christersen D, Preisser J, Miller W, Ibrahim M. Accessing a new

medication in Germany: a novel approach to assess a health insurance-related barrier. *Ann Epidemiol* 2005; **15**(10): 756–61.

368 Krobot K, Miller W, Kaufman J, Christ D. The disparity in access to new medication by type of health insurance: lessons from Germany. *Med Care* 2004; **42**(5): 487–91.

369 Krobot K, Miller W, Kaufman J, Christensen D, Preisser J, Ibrahim M. Quantifying delay in access to new medical treatment: an application of risk advancement period methodology. *Epidemiology* 2004; **15**(2): 202–7.

370 Atella V, Kopinska JA. The impact of cost-sharing schemes on drug compliance in Italy: evidence based on quantile regression. *Int J Public Health* 2014; **59**(2): 329–39.

371 Berger A, Sadosky A, Dukes E, Martin S, Edelsberg J, Oster G. Characteristics and patterns of healthcare utilization of patients with fibromyalgia in general practitioner settings in Germany. *Curr Med Res Opin* 2008; **24**(9): 2489–99.

372 Poole C, Tetlow T, McEwan P, Holmes P, Currie C. The prescription cost of managing people with type 1 and type 2 diabetes following initiation of treatment with either insulin glargine or insulin detemir in routine general practice in the UK: a retrospective database analysis. *Curr Med Res Opin* 2007; **8**: 41–8.

373 Rathmann W, Haastert B, Giani G. Drug prescriptions and costs in the treatment of diabetic polyneuropathy. *Dtsch Med Wschr* 1999; **124**: 681–6.

374 Icks A, Haastert B, Giani G, Rathmann W. Incremental prescription and drug costs during the years preceding diabetes diagnosis in primary care practices in Germany. *Exp Clin Endocrinol Diabetes* 2006; **114**: 348–55.

375 Gulliford M, Latinovic R, Charlton J. Diabetes diagnosis, resource utilization, and health outcomes. *Am J Manag Care* 2008; **14**(1): 32–8.

376 Mazzaglia G, Yurgin N, Boye KS, *et al*. Prevalence and antihyperglycemic prescribing trends for patients with type 2 diabetes in Italy: a 4-year retrospective study from national primary care data. *Pharmacol Res* 2008; **57**(5): 358–63.

377 Blak B, Mullins C, Shaya F, Simoni-Wastila L, Cooke C, Weir M. Prescribing trends and drug budget impact of the ARBs in the UK. *Value Health* 2009; **12**(2): 302–8.

378 Hughes D, McGuire A. The direct costs to the NHS of discontinuing and switching prescriptions for hypertension. *J Human Hypertens* 1998; **12**(8): 533–7.

379 Tate A, Nicholson A, Cassell J. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010; **102**(6): 947–51.

380 Keane MG, Horsfall LJ, Rait G, Pereira SP. Sociodemographic trends in the incidence of pancreatic and biliary tract cancer in UK primary care. *PloS One* 2014; **9**(9): e108498.

381 Garcia-Gil M, Elorza JM, Banque M, *et al*. Linking of primary care records to census data to study the association between socioeconomic status and cancer incidence in Southern Europe: a nation-wide ecological study. *PloS One* 2014; **9**(10): e109706.

382 Hamer HM, Kostev K. Sociodemographic disparities in administration of antiepileptic drugs to adults with epilepsy in Germany: a retrospective, database study of drug prescriptions. *CNS Drugs* 2014; **28**(8): 753–9.

383 Conrad N, Judge A, Tran J, *et al*. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet* 2018; **391**(10120): 572–80.

384 Klop C, van Staa TP, Cooper C, Harvey NC, de Vries F. The epidemiology of mortality after fracture in England: variation by age,

sex, time, geographic location, and ethnicity. *Osteoporos Int* 2017; **28**(1): 161–8.

385 Busby J, Schroeder K, Woltersdorf W, *et al*. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. *Br J Gen Pract* 2013; **63**(609): e256–66.

386 van der Velde RY, Wyers CE, Teesselink E, *et al*. Trends in oral anti-osteoporosis drug prescription in the United Kingdom between 1990 and 2012: variation by age, sex, geographic location and ethnicity. *Bone* 2017; **94**: 50–5.

387 Vermeer-de Bondt PE, Schoffelen T, Vanrolleghem AM, *et al*. Coverage of the 2011 Q fever vaccination campaign in the Netherlands, using retrospective population-based prevalence estimation of cardiovascular risk-conditions for chronic Q fever. *PloS One* 2015; **10**(4): e0123570.

388 Voordouw AC, Sturkenboom MC, Dieleman JP, *et al*. Annual revaccination against influenza and mortality risk in community-dwelling elderly persons. *JAMA* 2004; **292**(17): 2089–95.

389 Joseph C, Goddard N, Gelb D. Influenza vaccine uptake and distribution in England and Wales using data from the General Practice Research Database, 1989/90–2003/04. *J Public Health* 2005; **27**(4): 371–7.

390 Aspinall SL, Smith KJ, Good CB, *et al*. Incremental cost effectiveness of pharmacist-managed erythropoiesis-stimulating agent clinics for non-dialysis-dependent chronic kidney disease patients. *Appl Health Econ Health Policy* 2013; **11**(6): 653–60.

391 Hynes DM, Stroupe KT, Greer JW, *et al*. Potential cost savings of erythropoietin administration in end-stage renal disease. *Am J Med* 2002; **112**(3): 169–75.

392 Zeliadt SB, Penson DF. Pharmacoeconomics of available treatment options for metastatic prostate cancer. *Pharmacoeconomics* 2007; **25**(4): 309–27.

393 Lich KH, Tian Y, Beadles CA, *et al*. Strategic planning to reduce the burden of stroke among veterans: using simulation modeling to inform decision making. *Stroke* 2014; **45**(7): 2078–84.

394 Dickinson KC, Sharma R, Duckart JP, Corson K, Gerrity MS, Dobscha SK. VA healthcare costs of a collaborative intervention for chronic pain in primary care. *Med Care* 2010; **48**(1): 38–44.

395 Lin GA, Redberg RF, Anderson HV, *et al*. Impact of changes in clinical practice guidelines on assessment of quality of care. *Med Care* 2010; **48**(8): 733–8.

396 Conlin PR, Colburn J, Aron D, Pries RM, Tschanz MP, Pogach L. Synopsis of the 2017 U.S. Department of Veterans Affairs/U.S. Department of Defense Clinical Practice Guideline: Management of Type 2 Diabetes Mellitus. *Ann Intern Med* 2017; **167**(9): 655–63.

14

Inpatient Databases

James A. Feinstein[1,2], Peter K. Lindenauer[3,4], Chris Feudtner[5,6], and Brian T. Fisher[6,7]

[1] *Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO, USA*
[2] *Adult & Child Health Consortium for Health Outcomes Research and Delivery Science, University of Colorado and Children's Hospital Colorado, Aurora, CO, USA*
[3] *Center for Quality of Care Research, Baystate Medical Center, Springfield, MA, USA*
[4] *Department of Medicine, Tufts University School of Medicine, Boston, MA, USA*
[5] *General Pediatrics, Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[6] *Center for Pediatric Clinical Effectiveness, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*
[7] *Pediatric Infectious Diseases, Department of Pediatrics and the Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA*

Nationally representative nonexperimental studies evaluating the clinical effectiveness, safety, and variation in inpatient therapeutic practices were previously few in number. The dearth of these studies was primarily a function of available data sources that provide ample and accurate clinical and medication data in the inpatient setting to allow sufficiently precise and unbiased estimates of treatment effects. Furthermore, the data source needs to allow the researcher to establish temporal sequence of in-hospital exposures and events to facilitate testing of hypotheses about cause–effect relationships. Fortunately, the past two decades have seen a virtual renaissance for the development of databases containing more detailed information on inpatient medication use and clinical outcomes than has historically been captured by previous administrative data. These contemporary databases routinely incorporate, albeit to varying degrees, inpatient laboratory and radiologic test results and electronic health record (EHR) clinical data. This compilation of data from multiple institutions and across multiple settings of care (e.g., office visits, emergency room visits, inpatient hospitalizations) into comprehensive augmented administrative databases has afforded researchers the opportunity to conduct rigorous observational studies to answer important pharmacoepidemiologic research questions.

In this chapter, we first provide a general overview of how inpatient databases can be used to conduct pharmacoepidemiologic research studies. We include descriptions of a variety of existing databases that capture inpatient encounters, including Cerner Health Facts (HF), Pediatric Health Information System (PHIS), Premier Healthcare Database (PHD), and Truven MarketScan. Next, we review the common strengths and weaknesses of using these types of inpatient databases to conduct pharmacoepidemiologic research. Then, we provide illustrative examples of published

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

pharmacoepidemiologic studies to demonstrate how these databases can be applied to particular types of research questions (temporal trends, clinical practice variation, adverse event evaluation and therapeutic monitoring, and comparative effectiveness studies). Finally, because inpatient databases continue to evolve rapidly with the inclusion of more information, we present several considerations for future research studies that utilize these valuable resources.

Description

The availability of inpatient and other databases can give pharmacoepidemiologists an opportunity to define and create datasets representing a large number of patients from a multitude of medical institutions across the United States (see also Chapters 12 and 13). In an appropriately performed study, these data can be a valuable primary or adjunctive resource in helping to define the effectiveness or safety of various medications or interventions. The availability of clinical and sometimes laboratory data elements in these databases affords the researcher versatility to answer pharmacoepidemiologic questions related to common or rare medical [1–3]. Data from medical encounters (including those recording in-hospital drug use and outcomes) for the same patient can often be linked longitudinally and the timing of a particular intervention or medication exposure can be ascertained. Formatting the data in such a way allows for the investigation of time variant medication exposures and to assess for outcomes and safety events that may be more distant in time to the actual exposure. As multiple hospitalizations for individual patients can be linked over time, these databases afford the opportunity to analyze both common and rare illnesses that may require frequent contact with the medical system (e.g., malignancy or autoimmune conditions). Furthermore, the continuous collection of data allows researchers to trend the impact of certain illnesses from year to year (e.g., methicillin-resistant *Staphylococcus aureus*). This ability to consider the chronological association of an exposure and outcome rather than just an ecological association has the potential to further strengthen the implication of such an association. Additionally, these databases often contain payment and/or insurance information that enables utilization and cost analyses to be performed. Finally, the datasets are often representative of a large geographic area (nationally or regional) and thus the data provide an opportunity to examine differences by region, including variations in treatment practices for a specific illness [4].

In the US, the common core data elements are similar between databases and the data are typically fully compliant with the Privacy Rule of the US Health Insurance Portability and Accountability Act (HIPAA). While variable names, values, and value labels may differ, each database generally contains the following subject-specific data elements:

- demographic data
- visit data (e.g., office visit, emergency room visit, or inpatient admission)
- ICD-9-CM and ICD-10-CM diagnosis and procedure codes associated with a specific instance of care
- detailed prescription data
- laboratory and radiologic imaging orders
- payment data (Table 14.1).

Each specific database may contain additional unique variables of interest.

A select set of augmented inpatient databases is summarized in alphabetical order below; some of these databases also offer data from other types of healthcare encounters, such as ambulatory or emergency department visits, which are discussed in more depth in Chapters 12 and 13. Table 14.2 provides a more granular comparison of these databases to further emphasize their particular strengths and limitations. Certainly,

Table 14.1 Examples of common data elements.

Data type	Selected data elements
Demographic	Date of birth Race Gender Admission date (or encounter date for outpatient encounters) Discharge date APR-DRG classification
Diagnoses	Discharge (or encounter) diagnosis based on ICD-9-CM or ICD-10-CM codes Order of discharge (or encounter) diagnoses
Pharmacy	Medication ordered Dose of medication Route of administration Day of administration (or date of fill for outpatients) Days supplied for outpatient encounters Pharmacy charge
Procedures	Procedures performed based on ICD-9-CM or ICD-10-CM codes Date of procedure
Supply	Supply ordered Day supply delivered Supply charge
Laboratory	Lab ordered Day lab delivered Lab charge *Availability of actual results varies by database*
Radiologic imaging	Imaging procedure ordered Utilization of contrast media Day imaging procedure was ordered *Availability of results varies by database*
Clinical	Clinical service provided Day service provided Charge associated with service

APR-DRG, all patient refined – diagnosis related groups.

this group of datasets is not exhaustive of all currently available databases containing inpatient data. However, these databases were chosen based on author familiarity and to provide the reader with a reasonable landscape of the different existing options.

Cerner Health Facts® Database

The Cerner Corporation (www.cerner.com) maintains an EHR-derived database called Health Facts (HF). HF is a nationally representative longitudinal database containing clinical, laboratory, and administrative data from more than 116 adult hospital systems across the United States that cumulatively include 84 million patients available for analysis [5]. The contributing hospitals are primarily urban teaching facilities representing a wide range of bed sizes. The database elements are populated from Cerner EHRs and include detailed laboratory results (e.g., source of specimens, timing of collection, and results with reference ranges). Hospitalized patients can be tracked longitudinally post discharge if they return for outpatient or inpatient encounters within the same health system.

Pediatric Health Information System Database

The Pediatric Health Information System (PHIS) (www.childrenshospitals.org) is a comparative pediatric administrative database containing clinical and financial data elements for over 18 million patient encounters cumulatively from 45 not-for-profit, tertiary children's hospitals in the US. The PHIS was created and is managed by the Children's Hospital Association (CHA) (formerly the Child Health Corporation of America, CHCA). CHA is owned cooperatively by free-standing, noncompeting children's hospitals within the US, most of which contribute to the PHIS database. Member hospitals represent most of the major metropolitan areas across the US. Data are updated on a quarterly basis and made available to each hospital through a web-based reporting tool. A deidentified medical record number allows a patient to be tracked across multiple admissions at a

Table 14.2 Comparison of selected inpatient databases that contain medication data.

Data source	Subject sample size (in millions)	Data setting	Underlying sample	Longitudinal patient linking	Laboratory & radiology results[a,b]	Electronic health record data[a,b]	Public insurance[a]	Disability[a]	Mortality[a]	External dataset linkage possible	HIPAA compliant	Relative cost of data	Specific comments
Cerner Health Facts (HF) [32]	84	OP, ED, IP	116 US hospitals	Yes	Yes	Yes	Medicaid Medicare	No	Yes, if death occurred in medical facility	No	Yes	$	Must be a member hospital to acquire data
Pediatric Health Information System (PHIS)	18	ED, IP	45 Free-standing US children's hospitals	Yes	No	No	Medicaid	No	Yes, if death occurred in medical facility	Yes	Yes	$	Pediatric-specific database; must be a member hospital to acquire data; limited outpatient data
Premier Healthcare Database (PHD) [33]	147	OP, ED, IP	700 US hospitals	Yes	Yes[c]	No	Medicaid Medicare	No	Yes, if death occurred in medical facility	No	Yes	$$	
Truven MarketScan [34]	240	OP, ED, IP	Enrollees of employer-based insurances	Yes	Yes	Yes	Medicare Advantage Medicaid	Yes	Yes, if death occurred in medical facility	Yes	Yes	$$$	Includes retail, mail order, specialty pharmacy claims

[a] Typically available for a significantly smaller subset of the core sample, of the order of 10%.
[b] Typically limited to specific set of values, e.g., oxygen saturation, blood pressure, respiratory rate, BMI; not available at all visits by same patient.
[c] PHD does collect microbiology specific data for a select subset of participating institutions.

particular hospital. Assuming that appropriate institutional review board approval has been obtained from the respective institution, a researcher can request that the unique identifiers for patients admitted to their PHIS-affiliated institution be descrambled into the original medical record number. This allows for the possibility of performing additional chart abstraction to supplement or validate the data contained in PHIS.

Premier Healthcare Database

Premier (www.premierinc.com) is a consortium of US not-for-profit hospitals and health systems, created and owned by an alliance of more than 200 hospital and health systems. The Premier Prospective Database (PHD) originated as a merger of Premier with American Health Alliance and Sun Alliance in 1997 and continues to be managed by Premier, Inc. The PHD contains information cumulatively for 147 million patients available for analysis. Since its inception, Premier has collected and managed data from more than 700 hospitals and in so doing captures approximately 20% of all hospitalizations in the United States. Premier member hospitals are located across the US and range in size and setting from small rural to large inner-city hospitals. The PHD represents hospitals that admit both children and adults but is not typically inclusive of free-standing children's hospitals.

Truven Health MarketScan Database

The Truven Health (www.truvenhealth.com) MarketScan Research Databases contain three core databases – Commercial, Medicare Supplemental, and Medicaid – that cumulatively represent 240 million patients available for analysis. Administrative and clinical data are collected for all locations of patient care: inpatient, outpatient, outpatient pharmacy, mail order, and specialty pharmacy. For a subset of the larger database (approximately 10–15%), the claims data can be linked to additional datasets such as productivity management; health risk assessment; dental, laboratory, and medical records; and hospital data – including inpatient medication use – at the deidentified patient level. For example, the Truven MarketScan Hospital Drug Database contains inpatient drug utilization data from hospital discharge records that can be linked to the core databases at the patient level, which allows researchers to study medication utilization across both inpatient and outpatient settings [6].

Strengths

Data Quality

Many existing inpatient databases have been collecting data for over a decade. Each of the databases has data quality assurance processes in place, ensuring that the data they contain meet quality standards. For example, PHIS data audits primarily check for valid entries (e.g., valid ICD-9-CM or ICD-10-CM diagnosis codes) and reasonable patient information (e.g., birth weight). Researchers utilizing these databases and/or contributing organizations typically receive quality and benchmarking reports based on internal and external data audits, including advice on the impact of missing data elements. Finally, although studies using these databases are performed retrospectively, the actual data are collected in a prospective fashion independent of the study itself. This eliminates some of the inherent biases commonly identified for traditional retrospective studies (e.g., recall bias, interview bias or data collection biases).

Data Accessibility

These HIPAA-compliant databases are typically available either commercially (HF, Truven MarketScan) or by membership as a contributing data partner (PHD, PHIS). The cost of data

can depend on the status of the data recipient (e.g., not for profit versus for profit) and the extent of information requested for the desired dataset. Depending on the data vendor, databases can be readily accessed via a virtual network, web interface, or with assistance from the database administrators. By performing defined queries to the parent database, a researcher can "collect" necessary data from years of admissions in a matter of hours to days. Once the data are obtained, additional data cleaning and manipulation are usually necessary to establish a dataset suitable for analysis. For example, the data can be easily transformed into a time-dependent format for survival analysis.

Weaknesses

When using these databases for pharmacoepidemiologic studies, a researcher must consider several common limitations. First, the generalizability of study findings will depend on the degree to which the institutions that contribute to the dataset differ from noncontributing institutions, in terms of clinical practice or patient case mix. For example, the PHIS may accurately reflect practices within a free-standing children's hospital, but may not be generalizable to the typical community hospital. Similarly, the Truven MarketScan databases may allow researchers to ask questions that were previously impossible to answer without laboratory and radiology results, but the findings may not be generalizable because the databases are built from a large, nonrandom convenience sample of insurance claims from large employers [6].

Second, the research questions that these datasets can address may be limited by the lack of certain data elements. For example, study designs could be improved if the results of laboratory or radiographic tests performed on particular days of interest were reliably available (e.g., within 30 days prior to the index date). Even for datasets that include these types of data, laboratory and radiology results are not necessarily available for all patients or for all encounters by the same patient. Similarly, information regarding the precise day within a hospitalization on which a specific diagnosis was made, timing of onset of a complication, or documentation of a patient's condition at the time of admission or discharge are not always available. While alternative approaches can be taken to overcome certain data ascertainment challenges (e.g., creation of a severity of illness metric composed of resource utilization) this is not always possible. The existence of these metrics would enable more sophisticated analysis and allow for the researcher to draw conclusions with stronger inferences.

Finally, the validity of findings may be compromised because of some unknown degree of misclassification regarding several aspects of patient classification.

- Disease status, since the accuracy of clinical diagnoses, encoded as ICD-9-CM and ICD-10-CM codes at the time of hospital discharge, cannot be readily validated across all settings of care, and these diagnostic codes used for billing may not reflect the comprehensive set of clinical diagnoses that were made for a given patient.
- Exposure status, since drug exposure data are based on billing for dispensed drugs, which most likely but not always were administered to the patient.
- Outcome status, for the same reason mentioned above regarding disease status, as well as the possibility of the outcome occurring after hospital discharge and the patient either being readmitted to another hospital or not hospitalized.

Particular Applications

The comprehensiveness of these augmented databases makes them attractive options for testing clinically important pharmacoepidemiologic hypotheses in adult and pediatric patients,

hypotheses that were previously limited by insufficient data. Pharmacoepidemiologists have already utilized these databases to publish important findings that consider the diagnosis of less common outcomes after a pharmaceutical exposure, temporal trends of a particular disease, descriptions in variations of therapeutic practices, reports on therapeutic or drug toxicity monitoring, and comparative effectiveness studies. Examples of both pediatric and adult studies for each of these categories are discussed below.

Temporal Trends

Venous thromboembolism

Venous thromboembolism (VTE), consisting of a blood clot in the medium or large vessel venous circulation, poses several health hazards, including potential subsequent pulmonary embolism, thrombophlebitis, venous stasis, and diminished central venous access. It was hypothesized that the epidemiology of VTE was changing, due on the one hand to increasing use of intravascular venous catheters in ill patients and on the other hand to increasing prevalence of obesity and use of oral contraceptive hormones. Each of these factors, by different mechanisms, predisposes to the formation of VTEs. Data regarding the incidence of pediatric VTE, however, were sparse, consisting of two studies in Canada and The Netherlands, with only three and two years of data from the 1990s. Thus, both studies lacked sufficient precision to detect temporal trends, should they have existed. Furthermore, standard treatment for VTE includes anticoagulation, which with the advent of fractionated low molecular weight heparin in the 1990s has shifted to some degree from warfarin to enoxaparin, but the degree of this therapeutic shift had not been measured.

A study using PHIS, consisting of 41 children's hospitals that contributed data continuously from January 2001 to December 2007, identified (among 2.9 million hospitalizations) 13 449 hospital admissions of 9936 patients, with one hospitalization associated with a VTE diagnosis and 1401 patients with recurrent VTE diagnoses across several hospitalizations [7]. The annual proportion of VTE-associated admissions between 2001 and 2007 rose from 34 to 58 cases per 10 000 admissions, a 70% increase ($P<0.001$). The upward trend was observed across the age spectrum (Figure 14.1A). During the same time period, the proportion of patients diagnosed with VTE who received enoxaparin rose from 29% to 49%, while the proportion receiving warfarin declined slightly from 11% to 10% ($P<0.001$ for both trends) (Figure 14.1B). These findings helped to focus attention on improving the prevention, diagnosis, and treatment of VTE in pediatric patients.

Clinical Practices Variation

Henoch–Schönlein Purpura (HSP)

The most common pediatric vasculitis is HSP, and up to 40% of children with HSP are hospitalized to manage acute disease manifestations such as severe pain, gastrointestinal bleeding, hypertension, or glomerulonephritis [8]. There is no standard approach to diagnose and treat HSP. Clinicians use various laboratory and radiographic tests to make the diagnosis, and the subsequent treatment may include combinations of corticosteroids, antihypertensives, nonsteroidal antiinflammatory drugs (NSAIDs) and opioids, or nothing at all.

In order to describe inpatient variation in practice for HSP, researchers assessed hospitalization records from 36 children's hospitals contained in the PHIS between 2000 and 2007 [3]. Despite controversy regarding the effectiveness of corticosteroids in the treatment of HSP, 56% of patients during an initial hospitalization with HSP received this class of drug, compared to 36% who received opioids, 35% nonsteroidal antiinflammatory drugs, and 11% antihypertensive drugs. Substantial variation in the use of these medications was evident across the hospitals (Figure 14.2), and this variation persisted despite adjustment for case-mix differences

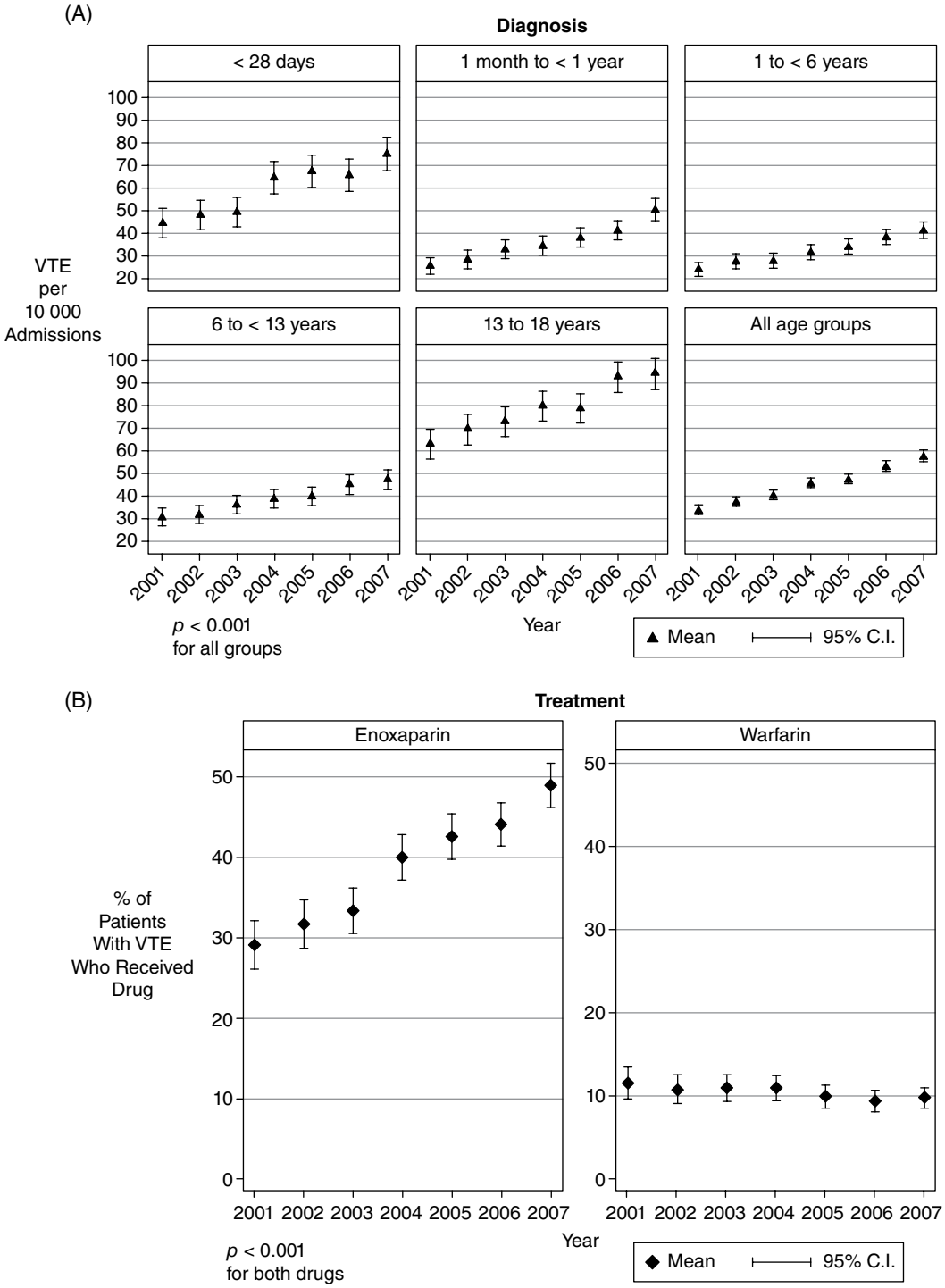


Figure 14.1 Temporal trends in (A) diagnosis and (B) treatment of VTE. *Source:* Adapted from Raffini *et al*. [7] with permission from the American Academy of Pediatrics.

among hospitals. These findings underscored the need for additional pharmacoepidemiologic research to enable the formulation of evidence-based practice guidelines.

Adverse Event Evaluation and Therapeutic Monitoring

Medication Use in Hospitalized Elderly Patients

Older adults tend to respond to certain medications differently from younger patients, with either reduced drug effectiveness or heightened susceptibility to adverse effects. Consequently, it is deemed best to avoid some drugs when treating geriatric patients. The Beers list, first developed in 1991 with multiple subsequent revisions, identifies such drugs to be avoided [9]. This list has been employed by the US Centers for Medicare and Medicaid Services and the US National Committee on Quality Assurance for regulatory and quality of care measurement purposes.

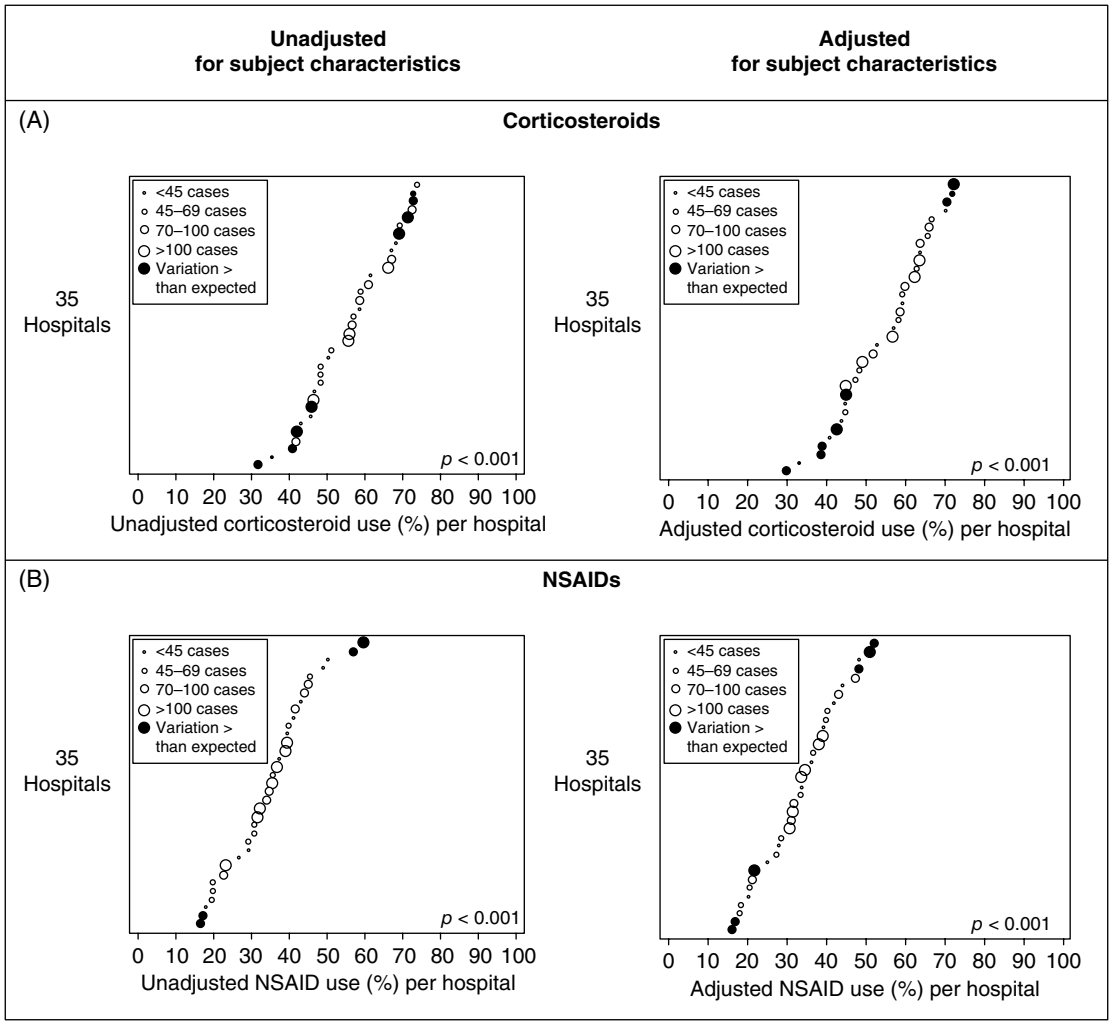


Figure 14.2 Variation in (A) corticosteroid and (B) NSAID use among hospitals in unadjusted and adjusted models. *Source:* Adapted from Weiss *et al.* [3] with permission from Elsevier.

Rothberg et al. used the PHD to identify 493 971 patients over 65 years of age (with a mean age of 78 years) cared for in 384 hospitals from 2002 to 2005 [10]. The study found that 49% of these inpatients were exposed to at least one potentially inappropriate medication (PIM, as defined by the 2002 Beers list), with 6% receiving three or more PIMs. The most commonly used PIMs were promethazine, diphenhydramine, propoxyphene, clonidine, amiodarone, and higher doses of lorazepam. Compared to internists, geriatricians were less likely to prescribe high-severity PIMs (adjusted odds ratio [AOR] 0.69; 95% confidence interval [CI] 0.61–0.78) as were hospitalists (AOR 0.90; 95% CI 0.84–0.96). However, cardiologists (AOR 1.32; 95% CI 1.28–1.36) and pulmonologists (AOR 1.10; 95% CI 1.05–1.15) were more likely to prescribe these high-severity PIMs. Of note, seven hospitals with more than 300 patients each had no PIMs dispensed to any of their elderly patients, suggesting a possible mechanism for future quality improvement efforts. Alternatively, targeted management of just three drugs (promethazine, diphenhydramine, propoxyphene) by hospital formularies or pharmacies would eliminate the use of PIMs in 24% of the geriatric patients.

This study example not only highlights the potential to identify areas for concern but also represents an opportunity to efficiently direct resources to effect a change.

Death and Cefepime Exposure

Cefepime is a fourth-generation cephalosporin antibiotic with broad-spectrum gram-positive and gram-negative activity. It is available as an intravenous formulation and is often used as empiric therapy for hospitalized patients with suspected bacteremia and sepsis. The administration of chemotherapy to children with cancer often renders them neutropenic with a high risk for bacterial infections. When a child becomes febrile during a period of neutropenia, it is recommended that they be admitted to hospital for initiation of broad-spectrum antibiotics [11]. In many instances, cefepime is the primary antibiotic utilized for these fever and neutropenia episodes. In 2007 a metaanalysis was published that questioned the safety of cefepime relative to other broad-spectrum antibiotics. The study pooled mostly adult randomized trials from various patient populations and found an increased risk of all-cause mortality in patients receiving cefepime compared to those receiving other beta-lactam antibiotics [12].

These data raised significant concern about the safety of cefepime. However, uncertainty existed regarding the true implications of the results from the metaanalysis for multiple reasons. First, the pooled studies included patient populations that were heterogeneous and included primarily adult studies; second, for each included study, mortality was not a primary endpoint, which called into question the completeness of this data point as the primary endpoint for the metaanalysis; lastly, a plausible mechanism for the increased risk for mortality secondary to cefepime was not identified. It was clear that additional data and analyses were needed to further evaluate the questioned association, especially among pediatric patients.

Therefore, using the PHIS database, a retrospective nationally representative homogeneous cohort of pediatric patients with acute myeloid leukemia (AML) was assembled. Children with AML were chosen for this analysis as they have frequent episodes of fever and neutropenia resulting in significant exposures to broad-spectrum antibiotics such as cefepime. Additionally, AML unfortunately carries a high mortality rate, thus establishing a cohort with a significant exposure to cefepime and more frequent rate of the outcome of interest (death). In total, 917 children in the PHIS database were found to have an ICD-9 code and chemotherapy receipt consistent with AML between 2002 and 2006. Table 14.3 displays the demographic characteristics of this PHIS-created cohort in comparison to the demographics of a cohort of patients with AML enrolled in a prospective chemotherapy trial sponsored by the Children's

Table 14.3 Comparison of demographic variables from a pediatric AML cohort created retrospectively in PHIS and another cohort assembled in a prospective chemotherapy trial [13,14].

	Retrospective PHIS AML cohort [21] (n=917)	Prospective AML chemotherapy trial [20] (n=492)
Sex		
Male (%)	513 (56%)	263 (53%)
Age		
Median (IQR)	9.2 (2.9 to 14.2)	9.6 (Not available)
0 to less than 2 yrs	186 (20.3%)	107 (22%)
2 to less than 16 yrs	603 (65.8%)	318 (65%)
Older than 16 yrs	128 (14.0%)	67 (14%)
Race		
White	649 (71%)	316 (64%)
Black	119 (14%)	42 (9%)

AML, acute myeloid leukemia; IQR, interquartile range; PHIS, Pediatric Health Information System.

Oncology Group. As the table illustrates, the two cohorts have similar distributions of gender, age, and race, establishing some external validity to the cohort created retrospectively via the PHIS [13].

After extracting all admission information for each patient for up to one year from diagnosis, a survival dataset was created. After adjusting, in a Cox regression model, for potential confounding factors such as age, gender, race, severity of illness, and variation in proportional hazards, there were no identified differences in all-cause in-hospital mortality in the year after treatment between patients recently exposed to cefepime versus those exposed to ceftazidime (hazard ratio [HR] 1.29; 95% CI 0.53–3.15), an antipseudomonal penicillin (HR 1.08; 95% CI 0.44–2.66), or carbapenem (HR 1.03; 95% CI 0.45–2.33) [14].

These results had an immediate impact by giving pediatric clinicians reassurance in the continued use of cefepime in their patient population. In fact, the data from this study were shared with the US Food and Drug Administration (FDA), which at the time was performing a review of the results of the initial metaanalysis. The FDA's own review and metaanalysis did not identify an increased risk of death associated with cefepime overall, nor did an additional pediatric-specific metaanalysis and systematic review [15].

Comparative Effectiveness

Antibiotics for Chronic Obstructive Pulmonary Disease

It has been estimated that as many as 24 million US residents suffer from chronic obstructive pulmonary disease (COPD) [16]. Acute exacerbations of COPD require inpatient care, resulting in as many as 600 000 hospital admissions annually, totaling an estimated $20 billion in direct costs each year [17]. Infection is a primary contributor for such exacerbations. Prior iterations of COPD treatment guidelines supported antibiotic administration at the time of admission in those COPD exacerbations that are associated with purulent sputum, an increase in sputum production, or an increase in dyspnea [18–20]. At the time, this recommendation for antibiotics was based on limited data from relatively small randomized trials, most of which were performed close to two decades previously [21].

A retrospective cohort of patients hospitalized between January 1, 2006 and December 31, 2007 for an acute exacerbation of COPD was

assembled using the PHD. Patients were included in the analysis if they were at least 40 years old, had a principal ICD-9-CM code for an acute exacerbation of COPD and did not have other infectious diagnoses (e.g., pneumonia, cellulitis). Using this cohort, the investigators compared the effectiveness of antibiotics in the first two days of admission to no antibiotics in those first two days [16]. The primary outcome for the analysis was a composite measure of progression to mechanical ventilation, in-hospital mortality, and readmission within 30 days of admission.

The comprehensive analyses included various multivariable models which utilized propensity scores (see Chapter 43) to balance measured baseline measured confounders that may have contributed to the clinicians' choice for starting or not starting antibiotics in the first two hospital days. Additionally, a logit link generalized estimating equation excluding antibiotic status was implemented to predict the risk of treatment failure so that the impact of antibiotics could be evaluated across three strata of predicted risk of treatment failure.

The cohort consisted of 84 621 patients, 79% of whom received antibiotics in the first two hospital days. In the propensity score and covariate adjusted model, the resultant odds ratio was 0.87 (95% CI 0.82–0.92) favoring early antibiotic administration. This apparent beneficial effect of early antibiotic use was most pronounced among those patients deemed to have the highest risk for treatment failure. Importantly, the presence of increased or purulent sputum production, as reflected by a ICD-9-CM procedure code for sputum testing on hospital day 1 or 2, did not alter the point estimate; similarly, the presence of severe dyspnea, as reflected by a ICD-9-CM procedure code for arterial blood gas measurement on day 1 or 2, did not alter the point estimate. This study identified a potentially important benefit for early antibiotic administration in the setting of COPD exacerbation. Furthermore, the results suggest that in contrast to the recommendations contained in prior iterations of the guidelines, antibiotics were associated with improved outcomes in patients requiring hospitalization.

Although this study cannot be considered equivalent to a large randomized controlled trial, the importance of its results to guiding clinical care should not be underestimated. This was an efficient and relatively inexpensive approach to establish a large cohort to analyze the utility of antibiotics for acute exacerbations of COPD. The statistical analysis was thorough and the results provide clinicians with reasonable estimates of the benefits for early antibiotic initiation in this group of patients.

Activated Protein C and Sepsis

Sepsis has been and will continue to be a major cause of inpatient morbidity and mortality. It has been estimated that close to 10% of all intensive care unit admissions are sepsis related, that sepsis-associated mortality ranges from 30% to 50%, and that the estimated annual cost for sepsis approaches $17 billion per year in the United States [22–24]. Given the high mortality of sepsis, researchers and clinicians have attempted to identify effective therapies to improve outcomes. One such therapy, human recombinant activated protein C (APC), has been evaluated in various prospective trials with mixed results [25–27]. Furthermore, these trials raised a concern about hemorrhagic complications from the APC therapy. Additional randomized controlled trials were being planned but those results will be several years in the making, leaving clinicians without further guidance on their current patients.

Therefore, a retrospective study using the PHD was performed to investigate the effectiveness of APC in reducing in-hospital mortality due to sepsis and, importantly, to report the rates of hemorrhagic complications [28]. The study included a cohort of patients admitted to one of 404 hospitals between June 1, 2004 and June 30, 2006. Patients were deemed to have sepsis if they had an ICD-9-CM code consistent

with sepsis, were admitted to an intensive care unit and received antibiotics and vasopressors within the first two days of admission.

The investigators identified 33749 patients meeting the study eligibility criteria, of which 4.7% received APC in the first two hospital days. A multivariable analysis including patient and hospital characteristics was utilized to define each patient's propensity to be given APC (see Chapter 47). Subsequently, a multivariable model comparing patients treated and not treated with APC matched by propensity score was performed, suggesting a benefit of APC on mortality (oods ratio [OR] 0.87; 95% CI 080–0.95). Interestingly, the rates of gastrointestinal bleeding, intracranial hemorrhage, and major transfusions were similar between those treated with APC and those not treated. While the results from this study do not negate the importance of previous and concurrently active randomized controlled trials, this well-done nonrandomized study provided both timely and relevant data regarding the treatment of sepsis in routine clinical practice.

The Future

Inpatient databases will continue to rapidly evolve given the ever-increasing types and volume of available clinical data. The size, structure, and complexity of these data will require researchers to employ creative new approaches and cutting-edge techniques from the fields of pharmacoepidemiology, biostatistics, and data science, including advanced approaches for addressing confounding in observational studies, methods for longitudinal data analysis, procedures for exploratory data mining and machine learning, and techniques for data visualization [29,30]. As analytic capabilities advance, pharmacoepidemiologists will have unprecedented opportunities to ask and answer new and innovative research questions across broad patient populations.

Several of the database options reviewed in this chapter already include internal linkages to certain types of novel patient data; additionally, the data in many of these databases can also be combined or merged with other novel data sources, such as a data source containing patient-specific data from emergency department or ambulatory clinic encounters, or about hospitals' processes of care. For example, a study of infection control practices in children's hospitals supplemented PHIS data with data regarding each hospital's level of use of alcohol-based hand hygiene gel (gathered by a survey). The resultant dataset was used to demonstrate a reduced odds of nosocomial gastrointestinal infections (AOR 0.64; 95% CI 0.49–0.85) among children cared for in hospitals where the hand gel was present [31]. Such data linkages will continue to increase in the era of improved capture of EHR data, patient-reported outcomes data, and, perhaps not so far off in the future, the results of pharmacogenomic testing.

Additionally, improved generalizability may result from analyzing the same pharmacoepidemiologic question using each of the different available databases. Because each of the discussed databases represents different medical institutions, and thus different patient populations, such an approach would help to prove the generalizability of certain outcomes or to identify important differences in outcomes that may be attributable to variability across the types of medical institutions. For example, the data from the PHIS and PHD can be combined to provide sampling coverage of both children's hospitals (as represented in the PHIS and to a much lesser extent in the PHD) and general hospitals (as represented only in the PHD), thereby presenting a more complete and accurate representation of the entire realm of pediatric hospital care. Likewise, analyzing multiple inpatient databases that include the entire range of commercially and publicly insured subjects will ensure more generalizable results that improve pharmacotherapy and enhance drug safety for the largest possible number of patients.

References

1 Hauptman PJ, Swindle J, Burroughs TE, Schnitzler MA. Resource utilization in patients hospitalized with heart failure: insights from a contemporary national hospital database. *Am Heart J* 2008; **155**(6): 978–85 e971.

2 Zaoutis T, Localio AR, Leckerman K, Saddlemire S, Bertoch D, Keren R. Prolonged intravenous therapy versus early transition to oral antimicrobial therapy for acute osteomyelitis in children. *Pediatrics* 2009; **123**(2): 636–42.

3 Weiss PF, Klink AJ, Hexem K, *et al*. Variation in inpatient therapy and diagnostic evaluation of children with Henoch Schonlein purpura. *J Pediatr* 2009; **155**(6): 812–18 e811.

4 Fisher BT, Chiller TM, Prasad PA, Beveridge M, Walsh TJ, Zaoutis TE. Hospitalizations for coccidioidomycosis at forty-one children's hospitals in the United States. *Pediatr Infect Dis J* 2010; **29**(3): 243–7.

5 DeShazo JP, Hoffman MA. A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample. *BMC Health Serv Res* 2015; **15**: 384.

6 Truven Health MarketScan Databases for life sciences researchers. https://truvenhealth.com/Portals/0/Assets/2017-MarketScan-Databases-Life-Sciences-Researchers-WP.pdf (accessed April 10, 2019).

7 Raffini L, Huang YS, Witmer C, Feudtner C. Dramatic increase in venous thromboembolism in children's hospitals in the United States from 2001 to 2007. *Pediatrics* 2009; **124**(4): 1001–8.

8 Yang YH, Hung CF, Hsu CR, *et al*. A nationwide survey on epidemiological characteristics of childhood Henoch–Schonlein purpura in Taiwan. *Rheumatology* 2005; **44**(5): 618–22.

9 American Geriatrics Society Beers Criteria Update Expert Panel. American Geriatrics Society updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc* 2012; **60**(4): 616–31.

10 Rothberg MB, Pekow PS, Liu F, *et al*. Potentially inappropriate medication use in hospitalized elders. *J Hosp Med* 2008; **3**(2): 91–102.

11 Hughes WT, Armstrong D, Bodey GP, *et al*. 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer. *Clin Infect Dis* 2002; **34**(6): 730–51.

12 Yahav D, Paul M, Fraser A, Sarid N, Leibovici L. Efficacy and safety of cefepime: a systematic review and meta-analysis. *Lancet Infect Dis* 2007; **7**(5): 338–48.

13 Sung L, Lange BJ, Gerbing RB, Alonzo TA, Feusner J. Microbiologically documented infections and infection-related mortality in children with acute myeloid leukemia. *Blood* 2007; **110**(10): 3532–9.

14 Fisher BT, Aplenc R, Localio R, Leckerman KH, Zaoutis TE. Cefepime and mortality in pediatric acute myelogenous leukemia: a retrospective cohort study. *Pediatr Infect Dis J* 2009; **28**(11): 971–5.

15 Kim PW, Wu YT, Cooper C, *et al*. Meta-analysis of a possible signal of increased mortality associated with cefepime use. *Clin Infect Dis* 2010; **51**(4): 381–9.

16 Rothberg MB, Pekow PS, Lahti M, Brody O, Skiest DJ, Lindenauer PK. Antibiotic therapy and treatment failure in patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. *JAMA* 2010; **303**(20): 2035–42.

17 Snow V, Lascher S, Mottur-Pilson C. Evidence base for management of acute exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med* 2001; **134**(7): 595–9.

18 National Collaborating Centre for Chronic Diseases. Chronic obstructive pulmonary disease. National clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. *Thorax* 2004; **59**(Suppl 1):1–232.

19 Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with

COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004; **23**(6): 932–46.
20 O'Donnell DE, Hernandez P, Kaplan A, *et al*. Canadian Thoracic Society recommendations for management of chronic obstructive pulmonary disease – 2008 update – highlights for primary care. *Can Respir J* 2008; **15**(Suppl A): 1A–8A.
21 Ram FS, Rodriguez-Roisin R, Granados-Navarrete A, Garcia-Aymerich J, Barnes NC. Antibiotics for exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2006; 2: CD004403.
22 Karlsson S, Varpula M, Ruokonen E, *et al*. Incidence, treatment, and outcome of severe sepsis in ICU-treated adults in Finland: the Finnsepsis study. *Intensive Care Med* 2007; **33**(3): 435–43.
23 Linde-Zwirble WT, Angus DC. Severe sepsis epidemiology: sampling, selection, and society. *Crit Care* 2004; **8**(4): 222–6.
24 Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 2003; **348**(16): 1546–54.
25 Abraham E, Laterre PF, Garg R, *et al*. Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med* 2005; **353**(13): 1332–41.
26 Bernard GR, Margolis BD, Shanies HM, *et al*. Extended evaluation of recombinant human activated protein C United States Trial (ENHANCE US): a single-arm, phase 3B, multicenter study of drotrecogin alfa (activated) in severe sepsis. *Chest* 2004; **125**(6): 2206–16.
27 Bernard GR, Vincent JL, Laterre PF, *et al*. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 2001; **344**(10): 699–709.
28 Lindenauer PK, Rothberg MB, Nathanson BH, Pekow PS, Steingrub JS. Activated protein C and hospital mortality in septic shock: a propensity-matched analysis. *Crit Care Med* 2010; **38**(4): 1101–7.
29 Bate A, Reynolds RF, Caubel P. The hope, hype and reality of Big Data for pharmacovigilance. *Ther Adv Drug Saf* 2018; **9**(1): 5–11.
30 Bate A. Guidance to reinforce the credibility of health care database studies and ensure their appropriate impact. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1013–17.
31 Zerr DM, Garrison MM, Allpress AL, Heath J, Christakis DA. Infection control policies and hospital-associated infections among surgical patients: variability and associations in a multicenter pediatric setting. *Pediatrics* 2005; **115**(4): e387–92.
32 Stefan MS, Priya A, Pekow PS, *et al*. The comparative effectiveness of noninvasive and invasive ventilation in patients with pneumonia. *J Crit Care* 2017; **43**: 190–6.
33 Premier. White Paper: Premier Hospital Database (PHD). https://learn.premierinc.com/i/790965-premier-healthcare-database-whitepaper/1 (accessed April 10, 2019).
34 Truven. Research Data and Analytic Tools. http://truvenhealth.com/markets/life-sciences/products/data-tools (accessed April 10, 2019).

Part IIIc

Studies Involving *Ad Hoc* Data Collection

15

Event Monitoring in the UK

Vicki Osborne and Saad A.W. Shakir

Drug Safety Research Unit, Southampton, UK

Pharmacovigilance involves the detection, assessment, understanding, and prevention of adverse effects attributable to medications, in addition to any other problems related to medication use [1]. Pharmacoepidemiology is an important source of evidence for pharmacovigilance, the goals of which are to enhance patient care and safety and to provide evidence for the effective assessment of benefit–risk profiles of medications [1]. The birth of modern-day pharmacovigilance arose from the consequences of thalidomide, which was licensed in the UK in 1958 and withdrawn in 1961 [2,3]. Globally, an estimated 8000–10 000 children were born with phocomelia to mothers who took thalidomide during pregnancy for morning sickness [4]. The length of time taken to identify the safety signal and establish causality was the stimulus for establishing pharmacovigilance systems that are still in use today to monitor for suspected adverse drug reactions (ADRs).

In 1964, the Committee on Safety of Drugs (CSD) in the UK was formed (which later became the Committee on Safety of Medicines [CSM] and is now the Commission on Human Medicines [CHM]), which subsequently established a national spontaneous reporting system, known as the Yellow Card system (see Chapter 10). This was followed by the Medicines Act which was signed into law in the UK in 1968, requiring efficacy, safety, and quality to be established with all marketed medications [5]. The reasons for monitoring postmarketing drug safety were summarized in 1971 in a report of the CSD [6]:

> No drug which is pharmacologically effective is entirely without hazard. The hazard may be insignificant or may be acceptable in relation to the drug's therapeutic action. Furthermore, not all hazards can be known before a drug is marketed; neither tests in animals nor clinical trials in patients will always reveal all the possible side effects of a drug. These may only be known when the drug has been administered to large numbers of patients over considerable periods of time.

Past experience has shown that unexpected hazards can occur with older medications as well as newly licensed drugs and this evidence provides support for the concept of "postmarketing surveillance." The limitations of premarketing studies for examining safety of medications have been outlined previously (see Chapter 1) and include, but are not limited to, sample size, duration of study, and the patient selection process. Therefore, there has been general agreement for more than 50 years on the importance of postmarketing adverse

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

event monitoring and postmarketing safety studies in providing complementary information on the clinically necessary understanding of the safety of a drug.

Study designs for postmarketing surveillance include both spontaneous reporting systems and other event monitoring methods. There are many advantages to spontaneous reporting systems, including the ability to identify rare ADRs, though limitations of such systems have been noted previously [7]. While spontaneous reporting methods are an important part of modern-day pharmacovigilance systems, nonpassive studies which actively monitor for events are also needed. Justification for active studies can be found through historical events such as in the case of practolol, which was licensed in 1970 by the CSD for the management of angina pectoris and cardiac dysrhythmias [8]. It was commercially a very successful drug but was associated with a rare condition known as oculomucocutaneous syndrome which was not detected by spontaneous reporting systems as the symptoms were not initially linked: rash, dry eyes, and sclerosing peritonitis [8,9]. The condition was eventually detected and the drug withdrawn in 1975 [8], although the five-year time period to detect this safety signal may have been avoided with the use of an active event monitoring system.

In June 1983, the CSM established a Working Party on Adverse Reactions which outlined a need for a prescription-based monitoring process to provide a method for monitoring new drugs.

The Postmarketing Drug Surveillance Research Unit was established in 1980 by Professor William Inman, to address the need for active drug monitoring [10,11]. Financial assistance for this endeavor was provided by the Office of the Chief Scientists of the Department of Health and Social Security; while it was initially formed as part of the Department of Medicine of the University of Southampton, the unit was reconstituted as a charitable trust in 1986 and the name was changed to the Drug Safety Research Unit (DSRU) which has remained to the present day. The DSRU event monitoring methodology is based on the fundamental concept of monitoring events regardless of relatedness to drug exposure, as first proposed by Finney in 1965 [10]. Over the years, this method has evolved from prescription event monitoring (PEM), to modified PEM (M-PEM) to the most recent design, specialist cohort event monitoring (SCEM). While PEM was retired by the DSRU in recent years, M-PEM and SCEM are still in use today.

The event monitoring designs used by the DSRU are the only national systems in the UK developed solely to monitor the utilization and safety of recently marketed medicines available to all primary care physicians (general practitioners [GPs]), besides the Yellow Card system. It is important to note that there are differences in the type of data collected in event monitoring compared with the Yellow Card system, the most important being that the majority of events reported with event monitoring will *not* be attributable to the drug (i.e., not adverse reactions) and should *not* be treated as spontaneous ADR reports. Nevertheless, both postapproval systems are able to generate hypotheses regarding safety signals. Event monitoring provides estimates of common to rare events while the Yellow Card scheme is able to detect signals of very rare events because of the size of the population being monitored. Thus, the Yellow Card spontaneous reporting system and event monitoring provide complementary information on hazards associated with medicines.

Description

Design and Source Data

Two different approaches are used to identify patients within M-PEM and SCEM studies, though the fundamental data collection design remains the same (see also subsection on Data Collection).

M-PEM Design and Data Source

Modified PEM is very similar to the original PEM method in that it uses an observational cohort design for active surveillance of targeted medicinal products in England. Products that are selected for study by M-PEM are usually new to the market, although established products can be monitored with specific justification; for example, a new indication or extending usage to a new population will usually require evidence from a postmarketing study to be submitted as part of the risk management plan (RMP). M-PEM utilizes the structure of the UK National Health Service (NHS), whereby all individuals are registered with a GP, who provides primary medical care and acts as a gateway to specialist and hospital care. Medical records (in paper and electronic form) are held for each individual within each general practice, are generally lifelong and are transferable among general practices when a patient moves to a new area. Medical records data include not only information obtained in primary care but information about all contacts with secondary and tertiary care, including letters from specialist clinics, hospital discharge summaries, results of laboratory and other investigations and information on GP-issued NHS prescriptions for the medicines the GP considers medically warranted.

Individuals are identified in M-PEM where they have received an NHS-issued prescription for the drug of interest. These prescriptions will be dispensed to patients at pharmacies with an NHS contract. Pharmacists are required to submit information on dispensed prescriptions for medications to a central prescription processing center within the NHS Business Services Authority (NHSBSA), formerly the Prescription Pricing Division (PPD). Reimbursement through the NHS can only be processed by following this procedure. Due to a long-standing agreement and via secure file transfer, the DSRU is provided with electronic copies of all those prescriptions issued throughout England for the drugs being monitored on a monthly basis (see also subsection on Ethics and Confidentiality). Since the NHSBSA only handles the remuneration and reimbursement to dispensing contractors across England, data are not available for Scotland, Wales, and Northern Ireland; however, where information is required from these areas, SCEM methodology can be utilized (see SCEM subsection).

The NHSBSA receives remuneration from the DSRU for providing prescription data. These data are reconciled with GP identifier records available from the NHS Organization Data Services (ODS), to obtain prescriber contact details and, with existing records on the DSRU customized database, to ascertain whether the data pertain to an existing eligible patient already within the DSRU database. It should be noted that all relevant prescriptions are collected, irrespective of whether they are a new or repeat course. Repeat prescriptions can be identified and separated from initial prescriptions for a medication for individual patients.

The DSRU arrangement with the NHSBSA operates for the necessary length of time required to collect a sufficient number of prescriptions to identify the required study sample size of patients. Since collection of dispensed prescription data usually begins immediately after the new drug has been launched, the eligible patient study population can be described either as an inception cohort (where the study drug is a new entity) or a new user cohort (where the drug under study might be a revised formulation and the patients may be regarded as "switchers" and exposed to the new formulation for the first time). In addition, as the data are sampled at national level, the cohort is representative of the population registered within the NHS in England.

SCEM Design and Data Source

To complement the already well-established technique of M-PEM, the DSRU developed

SCEM in response to the need to provide safety studies in secondary care and other specialist settings. SCEM methods allow the collection of exposure and outcome data on patients identified in other healthcare settings, such as hospitals and secondary care settings. By capturing data in specialist care through SCEM studies, safety data is collected on those who may be more complex in terms of underlying disease, co-morbidities, and concomitant medications than in the general disease population, as seen in M-PEM. SCEM methods differ from M-PEM methods in that patients are not identified via dispensed prescription data.

The DSRU works collaboratively with the UK National Institute of Health Research (NIHR) Clinical Research Networks (CRNs), which are associated with undertaking research with health service providers (trusts) at a national level. Prescribers are identified through these networks and patients are subsequently identified by these specialist prescribers and their informed consent is obtained for participation. Research personnel are used to aid with recruitment of prescribers to the study, although only prescribers have any contact with patients. As with M-PEM, data are sampled at national level and research personnel can be used to recruit in all areas of the UK, where required. In addition, the distribution of prescribers and patients is examined in further detail to ensure representativeness. Recruitment for a study continues until the desired sample size is achieved. A comparison of the M-PEM and SCEM study designs is provided in Figure 15.1.

Ethics and Confidentiality

Both M-PEM and SCEM studies are conducted according to national and international guidelines for ethical conduct of research involving human subjects [12–14]. Following the principles of good practice [15,16], a full research protocol is written for each study to monitor and research the safety of medicines. Where these protocols describe a postauthorization safety study that forms part of the EU RMP, European Medicines Agency guidelines are adhered to [17]. For M-PEM studies, under Section 251 of the NHS Act 2006, the DSRU has received support from the Ethics and Confidentiality Committee of the National Information Governance Board to gain access to and process patient identifiable information without consent for the purposes of medical research. In contrast, for SCEM studies, informed consent is obtained from each patient

	M-PEM Study	**SCEM Study**
Prescribers	Primary Care	Specialist Care
Design	Observational cohort	Observational cohort
Treatment	Naturalistic (non-interventional)	Naturalistic (non-interventional)
Period of Observation	Longer period (e.g. 12 months)	Shorter period (e.g. 12 weeks)
Ethics	No patient consent required	Patient consent required
Risk	Less likely to capture high risk patients	More likely to capture high risk patients

Figure 15.1 Comparison of M-PEM and SCEM study designs.

prior to participation in the study. A single UK-wide ethical opinion is applied for from the National Research Ethics Service and local Research and Development (R&D) approval is also required at each participating NHS trust. Considerable care is taken to preserve the confidentiality of patient data and the DSRU databases are fully protected. Patient information security is assured through strict measures guided by DSRU policies. Furthermore, highly confidential patient data (name and address) supplied by the NHSBSA for M-PEM studies are made anonymous through use of a unique study identifier code assigned by the DSRU and separately one supplied by the GP on the questionnaire at the point of return. These codes are used for any subsequent correspondence.

Data Collection

Process for M-PEM and SCEM

Relevant study data are currently collected via a manual process (as summarized in Figure 15.2).

The process for data collection in M-PEM and SCEM studies is usually the same, though in SCEM studies there is often a questionnaire sent at baseline in addition to the main study questionnaire. For both study designs, after a protocol-specified time interval (range 3–12 months) from the date of the first prescription for each eligible patient, questionnaires are sent to prescribers. For M-PEM studies, these questionnaires are sent by surface mail in monthly batches according to the chronological order of prescription issue date to those GPs who prescribed the newly marketed medicine, continuing until the target sample size is achieved. Prepaid envelopes are provided for return by surface mail. For SCEM studies, questionnaires are usually printed by the prescriber directly from the study website and returned via prepaid surface mail. However, the process is now in place for data to be entered electronically into a dedicated website by doctors in SCEM studies.

Historically, PEM questionnaires were commonly referred to as "Green Forms" because of their color. These were intended to be simple in order to expedite data collection in order to enhance surveillance and encourage response in the interest of drug safety, especially given there was no remuneration to respondents. Given that M-PEM and SCEM studies provide remuneration to prescribers, these questionnaires are more detailed than the original Green Forms and are no longer green in color. The number of M-PEM study questionnaires sent on a monthly basis to each GP is limited (maximum of four per GP), but questionnaires for patients excluded because of this rule are subsequently included in the following month. As illustrated

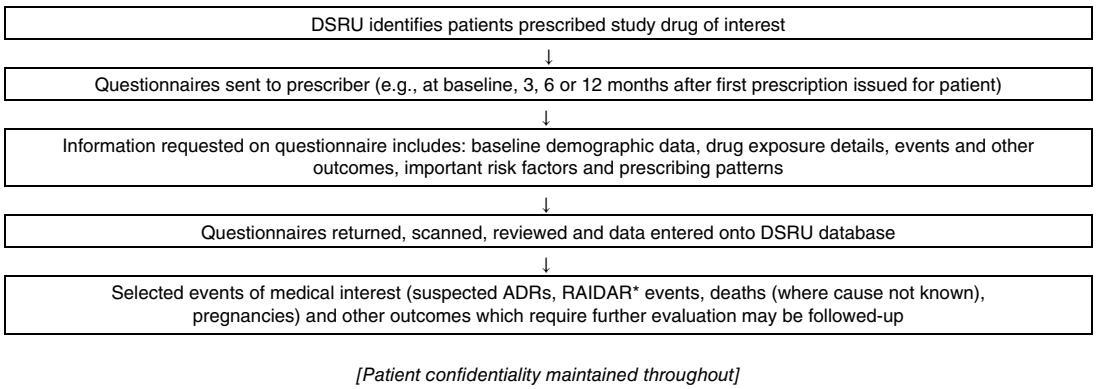


Figure 15.2 The event monitoring process. *RAIDAR, Rare and Iatrogenic Adverse Reactions.

Box 15.1 Definition of an event in event monitoring

Any new diagnosis, any reason for referral to a consultant or admission to hospital, any unexpected deterioration (or improvement) in a concurrent illness, any suspected drug reaction, any alteration of clinical importance in laboratory values, or any other complaint that was considered of sufficient importance to enter into the patient's notes.

in Figure 15.3, the standard Green Form PEM questionnaires were used to request data on patient demographics (age, gender), indication for treatment, prescribing information (dose and duration), reasons for stopping (if stopped), details of all significant events (for definition see Box 15.1) that have been recorded in patients' medical records during a specific time period after starting the PEM study drug, and cause(s) of death if applicable.

Questionnaire Design

In parallel with developments in the field of pharmacoepidemiology in general, event monitoring questionnaires have evolved over time to extend the range of data that could be collected on drug utilization patterns as well as important risk factors for selected outcomes of interest. In recent years, further expansion and modification of the nature and type of information requested on questionnaires for the whole cohort have led to the adoption of more complex forms; these are currently in use for both M-PEM and SCEM studies. Customized questionnaires have been developed to permit a wider exploration of more specific safety issues through collection of relevant information, while the underlying process remains the same.

Figure 15.4 illustrates the first page of the M-PEM questionnaire for lumiracoxib. Such customized questionnaires are designed to accommodate the increasing need for supplementary information which is relevant to a number of outcomes. Specifically, important identified and potential risks from the RMP, in addition to important missing information, can be explored. As a result, M-PEM and SCEM studies can explore a range of research questions, such as safety in special populations, drug utilization, and targeted surveillance of specific safety concerns. Because of the increased complexity of event monitoring questionnaires, prescribers receive remuneration for returning completed customized forms. Completion of these forms by prescribers remains voluntary.

Supplemental Information

During the course of any study, selected medical events (as described in Box 15.2) and other outcomes undergo preliminary evaluation for purposes of summarizing common or unusual features/manifestations, clinical course, and prognosis of conditions. Supplemental information may be sought from prescribers using targeted questionnaire(s), where such information is not provided on the main study questionnaires. For example, where a cerebrovascular accident (CVA) is reported, a specially designed questionnaire can be sent to obtain detailed information on the type of CVA, lab results, and risk factors for the event. Such questionnaires are sent within weeks of the initial review but in some cases, where an objective of a study might be to monitor events with a long latency, a lag period may be introduced, for example 12 months from the date of first occurrence of the event of interest, such as androgenic manifestations with testosterone use in women. However, the lag period can be increased significantly for studying delayed outcomes, such as reports of cancer. Remuneration is paid to physicians in appreciation of the time taken to complete these supplementary questionnaires.

Box 15.3 lists the medically serious events that have been associated with the use of medicines, as compiled by the DSRU. Cases of these events routinely undergo further evaluation (see

PLEASE REMOVE THIS SECTION OF THE FORM SO THAT THE BOTTOM HALF BECOMES ANONYMOUS AND KEEP FOR FUTURE REFRENCE IN THE EVENT OF FOLLOW-UP

DRUG SAFETY RESEARCH UNIT
PRESCRIPTION EVENT MONITORING

MEDICAL - IN CONFIDENCE

DR.

WOULD YOU PLEASE COMPLETE THIS QUESTIONNAIRE FOR:

WHO WAS PRESCRIBED

ON

Saad A W Shakir, FRCP (Glas & Ed),FFPM,MRCGP.
Bursledon Hall,
Southampton SO31 1AA
Telephone: (023) 80408600

We collect EVENT data

An EVENT is any new diagnosis, any reason for referral to a consultant or admission to hospital, any unexpected deterioration (or improvement) in a concurrent illness, any suspected drug reaction, any alteration of clinical importance in laboratory values or any other complaint which was considered of sufficient importance to enter in the patient's notes.
Example: A broken leg is an EVENT.

Please indicate if you suspect an EVENT to be an adverse reaction to a drug.

These studies are conducted in accordance with the results of authorative discussions and the International Ethical Guidelines for Biomedical Research Involving Human Subjects prepared by the Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO), Geneva 1993. The method of study also complies with the Guidelines on the practice of Ethics Committees in Medical Research Involving Human Subjects, as issued by the Royal College of Physicians of London (August 1996)

Your identification code for this patient*

Ref:

The DSRU is advised by the Drug Safety Research Trust, a registered independent charity (No. 327206). The unit operates in association with the University of Southampton.
Trustees: Dr John J Ferguson, Professor Sir Charles F George MD FRCP, Sir Gordon Higginson DL FEng, Professor Stephen T Holgate MD DSc FRCP, Professor Sir Michael D Rawlins MD FRCP FFPM, Dr Richard Tiner, Professor M P Vessey CBE MD FRCP FFPHM FRS.

PLEASE RETURN THIS HALF OF THE FORM Ref:

Your identification code for this patient*	Was the drug effective? Yes☐ No☐ Don't Know☐
Sex: Age at start of treatment:	Has the drug been stopped? Yes☐ No☐ Don't Know☐
Indication for prescribing	If 'YES' reason for stopping
Drug start date ___/___/___ Dose mg/day Date of last prescription ___/___/___	Drug stop date ___/___/___

Event Date	Dose mg/day	**Events while taking this drug** If none, please tick box ☐

Event Date	**Events after stopping this drug** If none, please tick box ☐

IF YOUR PATIENT HAS DIED: Date of death___/___/___
Certified cause of death 1a_____
Underlying cause_____

	Yes	No	Don't know
1. Prior to taking celecoxib was there any history of dyspeptic symptoms or other upper gastrointestinal conditions?	☐	☐	☐
2. Were any NSAIDs prescribed in the 3 months prior to celecoxib?	☐	☐	☐
3. Were any of the following drugs prescribed during treatment with celecoxib?	☐	☐	☐

	Yes	No	Don't Know		Yes	No	Don't Know
NSAIDs	☐	☐	☐	H2 antagonists/Proton pump inhibitors	☐	☐	☐
Misoprostol	☐	☐	☐	Antacids	☐	☐	☐
Aspirin	☐	☐	☐	Anticoagulants/Antiplatelet agents	☐	☐	☐

IMPORTANT: PLEASE INDICATE ANY EVENTS REPORTED TO CSM OR MANUFACTURER
This enables YOU to identify the patient in any future correspondence concerning this report

Figure 15.3 Green Form for the PEM study on Celebrex (celecoxib).

DSRU Reference: «srefCode»
Patient Year of Birth: «YOB»
Patient Sex: «sSexDescription»
Start Date «startdate»

Professor Saad Shakir
Bursledon Hall, Blundell Lane
Southampton, SO31 1AA
Telephone: (023) 8040 8600
Fax: (023) 8040 8609

LUMIRACOXIB (PREXIGE®) MODIFIED PEM STUDY
MEDICAL IN CONFIDENCE

PATIENT IDENTIFICATION CODE

Please provide an identification code that is only recognisable by you and can be used in the case of future correspondence about your patient if necessary.

Your Patent ID:

PRESCRIBING DATA

1. According to PPD data, this patient was first prescribed lumiracoxib on the date given above. Please confirm below:

a) Start date: ___/___/___ b) Dose at start : .....mg/day

c) Sex: M ☐ F ☐ d) YOB.....

2. Please confirm this patient is still registered with your practice

Yes ☐ ➧ *Go to* ***Question 3*** *and continue to complete questionnaire*

No ☐ ➧ *Left practice?* **Yes** ☐ ➧ *Please* ***return the questionnaire*** *in the* ***FREEPOST*** *envelope*

➧ *Or Died?* **Yes** ☐ ➧ *Please complete questionnaire if possible and provide details of death -* ***Question 13*** *overleaf*

3. Please specify the clinical prescribing indication(s) for lumiracoxib (tick all that apply)

Osteoarthritis (OA): OA Knee ☐ OA Hip ☐ OA Other ☐
Acute pain relief from surgery: Orthopaedic ☐ Dental ☐
Primary Dysmenorrhoea ☐

Other: please specify.....

4. Please indicate the reason for prescribing lumiracoxib for the above indications: (tick all that apply)

Prescriber decision ☐ Patient preference ☐
Initiated in secondary care ☐ Prescribing formulary ☐
Non-response to previous NSAID:
COX-2 selective inhibitor ☐ Other NSAID ☐
Intolerance to previous NSAID:
COX-2 selective inhibitor ☐ Other NSAID ☐

Other: please specify.....
.....

5. Please provide the following data regarding stopping treatment with lumiracoxib

Has treatment stopped? Yes ☐ No ☐ DK ☐
if Yes please provide:
a) Stop date ___/___/___ and/or last prescription date ___/___/___

Reason for stopping:.....
.....

PATIENT DATA

6. Please specify your patient's ethnicity, if known:

Caucasian ☐ Asian (Indian sub-cont) ☐ Hispanic ☐
Black ☐ Asian (China/Japan) ☐

Other.....

7. At the time of starting lumiracoxib, please specify this patient's status regarding general health:

Body Mass Index.....kg/m² DK ☐
Smoking status: Current ☐ Past (ex) ☐ Never ☐ DK ☐
Alcohol consumption: Occasional ☐ Excessive* ☐ Never ☐ DK ☐
*Male > 21 units/week; Female > 14 units/week

8. Prior to starting lumiracoxib did the patient have a history of the following: (tick all that apply)

	Yes	No	DK
Hypersensitivity reactions (e.g. angio-oedema/urticaria)	☐	☐	☐
Serious skin reactions (e.g. erythema multiforme)	☐	☐	☐
Excessive* alcohol use	☐	☐	☐

*Male > 21 units/week; Female > 14 units/week

9. At the time of starting lumiracoxib did the patient have any of the following: (tick all that apply)

	Yes	No	DK
Dyspeptic or other upper gastrointestinal conditions	☐	☐	☐
Cardiovascular disease (e.g. hypertension)	☐	☐	☐
Coronary disease (e.g. arrhythmia/infarct/cardiac failure)	☐	☐	☐
Cerebrovascular disease (e.g. CVA/TIA)	☐	☐	☐
Abnormal renal function/renal disease	☐	☐	☐
Diabetes	☐	☐	☐
Rheumatoid arthritis	☐	☐	☐
Abnormal liver function/hepatic disease*	☐	☐	☐

*Because of recent prescribing restrictions for lumiracoxib, we are collecting additional information on liver function tests and relevant risk factors **(see q14 and 15)***

10. Was this patient taking any of the following medications in the periods indicated: (tick all that apply)

	In 3 months prior to starting			During treatment		
	Yes	No	DK	Yes	No	DK
Antacids	☐	☐	☐	☐	☐	☐
Histamine $_2$ Antagonists / PPIs	☐	☐	☐	☐	☐	☐
Misoprostol	☐	☐	☐	☐	☐	☐
Aspirin (analgesia >300mg od)	☐	☐	☐	☐	☐	☐
Aspirin (cardioprotection <300mg od)	☐	☐	☐	☐	☐	☐
Other Anticoagulants / Antiplatelet agents	☐	☐	☐	☐	☐	☐
Non-selective NSAIDs	☐	☐	☐	☐	☐	☐
COX-2 selective NSAIDs (inc meloxicam)	☐	☐	☐	☐	☐	☐
Corticosteroids (oral/systemic)	☐	☐	☐	☐	☐	☐
Antidepressants (SSRIs and similar)	☐	☐	☐	☐	☐	☐
Antihypertensives	☐	☐	☐	☐	☐	☐
Statins or Fibrates	☐	☐	☐	☐	☐	☐
Sex Hormones (HRT/OC)	☐	☐	☐	☐	☐	☐

Figure 15.4 Questionnaire for the M-PEM study on Prexige (lumiracoxib).

DSRU Reference: «srefCode»

EVENT DATA

An EVENT is any new diagnosis, any reason for referral to a consultant or admission to hospital, any unexpected deterioration (or improvement) in a concurrent illness, any suspected drug reaction, any alteration of clinical importance in laboratory values or any other complaint which was considered of sufficient importance to enter in the patient's notes. Example: a broken leg is an EVENT.

IMPORTANT: please indicate any events reported to Commission on Human Medicines (CHM) [previously known as Committee on Safety of Medicines (CSM)]or manufacturer

11. Please list any events after starting treatment with lumiracoxib?

None ☐

Event date	Dose (mg/ day)	Event

12. Please list any events within 3 months after stopping lumiracoxib?

None ☐

Event date	Event

13. If your patient has died please provide:

Date of death:
Cause of death as recorded on death certificate:
1a
1b
2

LIVER FUNCTION

14. On starting lumiracoxib, please specify if this patient had any specific risk factors for abnormal liver function/hepatic abnormality?

	Yes	No	DK
Medical conditions			
Malignancy	☐	☐	☐
Autoimmune Disease	☐	☐	☐
Chronic Viral Infection (e.g Hepatitis B/C, EBV)	☐	☐	☐
Iron or copper overload	☐	☐	☐
Other conditions:	☐	☐	☐

If Yes to Other conditions, please specify:.....

.....

.....

.....

Concomitant medications

Paracetamol	☐	☐	☐
DMARDS (e.g methotrexate)	☐	☐	☐
Biologics (e.g beta-interferon)	☐	☐	☐
Other medications (prescribed, OTC, herbal products:	☐	☐	☐

If Yes to Other medications, please specify:.....

.....

.....

.....

15. Please provide liver function lab test values in the table below / send copies of reports as applicable on starting lumiracoxib and during treatment.

***NB** Please include **normal ranges** as they are vital for interpretation.*

	At start*	During treatment			After stopping	Normal range (Units)
		(1)	(2)	(3)**		
Date						
Bilirubin						
AST						
ALT						
GGT						
Alk Phos						
Albumin						

**nearest to the lumiracoxib start date*

*** If test results reported on more than three occasions during treatment please provide on a separate sheet as appropriate*

Thank you for your co-operation.
Your time and effort is greatly appreciated.

Please return in the FREEPOST envelope provided.

Figure 15.4 (Continued)

section on Qualitative Evaluation of Important and Medically Important Adverse Events). All pregnancies reported during treatment or within three months of stopping the drug are followed up using a supplementary questionnaire to determine the outcome of the pregnancy. All reported deaths for which no cause is specified are also followed up to try to establish the cause of death.

Data Processing

Each event monitoring questionnaire returned for each patient is scanned onto the database and electronic copies are reviewed by the DSRU research fellow monitoring the study. This initial review aims to identify possible serious ADRs or events requiring action, such as external communications or expedited follow-up. In accordance with DSRU procedures, any events of interest or deaths reported on questionnaires require further review by a clinical research fellow who will determine if supplementary information is required. It should be noted that event monitoring is designed to capture all events, regardless of attribution, unlike spontaneous reporting systems which collect events for which there is often an inherent assumption of a causal relationship with the treatment in the mind of the reporter. These events may be expected or unexpected and may be either serious or nonserious.

For each patient, trained coding staff prepare a computerized, longitudinal, chronological record of demographic, exposure, and outcome data associated with starting the study drug. All events reported on questionnaires are coded onto a DSRU database using MedDRA® [18]. This hierarchical dictionary, which is arranged by system-organ class, groups associated lowest level terms (terminology used by the prescribing physician) under preferred terms; similarly, related preferred terms are grouped under further broader terms (high-level term and higher level group term) [18]. Selected attributes can be linked to selected data; for example, an event can be flagged as a suspected ADR if the GP specified that the event was attributable to a drug (either the study drug or another drug taken during the study observation period). Other examples include if the event had a fatal outcome or if the event was a reason for stopping.

Good clinical data management is a high priority. The DSRU has a set of rules and processes associated with the conduct of all studies which undergo regular review. Data quality is assured through a number of methods based on error prevention, data monitoring, data cleaning, and documentation. For example, data cleaning is undertaken to screen for errors, missing values, or extreme values and to diagnose their cause.

Sample Size and Duration

As summarized in Chapter 4, the ability to detect an adverse event in a cohort study is dependent on the expected incidence rate of the adverse event in those exposed to the drug, the background rate in those not exposed to the drug, and the total number of patients. In original PEM studies, the sample size of 10 000 exposed patients was driven by PEM's original objective to bridge the gap between randomized trials and spontaneous reporting regarding sensitivity to rare and uncommon events that can be achieved by including a larger sample size than premarketing studies. Based on the general 'rule of 3' (see Chapter 4), it follows that the larger the sample size, the rarer the event that can be detected. A sample size of 10 000 patients allows one to be 95% certain that any events not observed occur less often than 1 in 3333 cases (incidence <0.0003) (see Chapter 4).

A sample size of 10 000 should allow for the detection of at least three cases of an adverse event, with 85% power, if the event occurs at a rate of at least one in 2000 patients (assuming the background rate is zero) [19]. If the background rate is known and there is an *a priori*

Box 15.2 Categories of events and outcomes which undergo further evaluation

Medically important[a] adverse events:
- Reported during premarketing development
- Reported during postmarketing in other countries (for products launched elsewhere before the UK)
- For the therapeutic class
- Previous undocumented medically important events considered to be possibly associated with the study drug during the study

Rare and Iatrogenic Adverse Reactions (RAIDAR) events (see Box 15.3)

Any other adverse events deemed to be of medical importance by the DSRU during the study

Specific outcomes associated with the study aims and objectives, for example aspects of prescribing, preexisting medical conditions or use of other medications immediately prior to or concurrently with the study drug which may be contraindicated, or which requires special warnings or precautions for use

[a] Defined as "events that may not be immediately life threatening or result in death or hospitalization, but may jeopardize the patient or require medical intervention to prevent serious sequelae."

Box 15.3 Rare serious adverse events that have been associated with the use of medicines

- Agranulocytosis
- Alveolitis
- Anemia, aplastic
- Anaphylaxis
- Angioneurotic edema
- Arrhythmia
- Bone marrow, abnormal
- Congenital abnormality
- Dermatitis, exfoliative
- Disseminated intravascular coagulation
- Erythema multiforme
- Erythroderma
- Guillain–Barré syndrome
- Hepatic failure
- Hepatitis
- Jaundice
- Leukopenia
- Multiorgan failure
- Nephritis
- Nephrotic syndrome
- Neuroleptic malignant syndrome
- Neutropenia
- Pancreatitis
- Pancytopenia
- Pseudomembranous colitis
- Renal failure, acute
- Retroperitoneal fibrosis
- Rhabdomyolysis
- Stevens–Johnson syndrome
- Sudden unexpected death
- Thrombocytopenia
- Torsade de pointes
- Toxic epidermal necrolysis
- Any event for which there is a positive rechallenge

This list is based on a similar list used by the Medicines and Healthcare products Regulatory Agency (MHRA), UK.

hypothesis of the effect size, then it is possible to analyze the statistical power of a study given a fixed sample size. For example, assuming 5% (two-sided test) significance, the power of a study based on 10 000 subjects to detect as statistically significant an increase in incidence from 0.1% to 0.2% would be 80% [20].

Because of the customized nature of M-PEM and SCEM studies, a specific sample size is calculated depending on the research question of interest, for which the outcomes are chosen and defined through internal DSRU scientific discussion as those which best reflect the research question. For example, if the study has been designed to test an increase in risk of an event between two study periods, then the sample size will depend on the background rate of the event, the estimated effect size of the adverse drug

event, the level of significance used (the fixed probability of wrongly rejecting the null hypothesis when no real difference exists; most commonly set at 5% – the type I error or false-positive result) and the desired power (the probability that the test will reject a null hypothesis that is false, i.e., that it will not make a type II error, usually taken to be 80%, which means a 20% chance of a type II error or false-negative result). The background rate and the expected adverse effect size are estimated from the published literature and clinical study data. For the majority of M-PEM and SCEM studies that have been undertaken to date, the sample size has been smaller than the 10 000 required for original PEM studies.

The duration of any study is dependent on the level of prescribing of the study drug by physicians in the UK; if the desired sample size has not been reached by a prespecified date then the study may either be extended or finalized, dependent on discussions with the marketing authorization holder and regulatory agencies. Interim analyses are usually undertaken at prespecified milestones (e.g., annually) or defined sample sizes (e.g., 2500 patients) and contacts are, whenever possible, maintained with the marketing authorization holder, so that the pharmaceutical companies (although the study is independent of them) can comply with the drug safety reporting procedures of the regulatory authorities.

The DSRU has completed 111 PEM studies to date with a median cohort size of 11 541 patients (interquartile range 8482–13 643). In addition, 13 M-PEM studies have been completed with a median cohort size of 4624 patients (interquartile range 1666–12 135). Two SCEM studies have also been completed to date and other M-PEM and SCEM studies are currently ongoing. A wide range of drugs have been studied using event monitoring, including agents to treat hypertension, angina, asthma and chronic obstructive pulmonary disease (COPD), diabetes, epilepsy, depression, schizophrenia, erectile dysfunction, and urinary incontinence. In addition, a number of nonsteroidal antiinflammatory drugs (including selective COX-2 inhibitors), several antibiotics, and antiviral agents have been studied.

Data Analysis

The primary objective of pharmacovigilance is signal detection and evaluation. Several methods are applied for signal detection in event monitoring, both qualitative and quantitative, not only to look for new unexpected adverse reactions but also for further information regarding expected drug–adverse event associations of interest that might affect the benefit–risk balance of a drug.

Qualitative Evaluation of Important and Medically Important Adverse Events

As described earlier, each questionnaire is evaluated for adverse events that may possibly be related to drug exposure. This qualitative evaluation by the DRSU research fellow takes into consideration a number of points (see Box 15.4). An example of a safety signal generated in event monitoring as a result of careful clinical evaluation is the visual field defects in patients receiving long-term treatment with the antiepileptic drug vigabatrin [21].

Because of the epidemiologic nature of the design of event monitoring studies, any inferences on drug-relatedness will be made on an aggregate basis at study milestones, such as when the interim and final reports are written. Such aggregate analyses can help formulate possible hypotheses, which then require further analytic study. Event monitoring is dynamic in nature and the types and nature of events evaluated may evolve during the course of the study, for example, following publications of case reports or regulatory concerns. An example is that of serious skin reactions and selective COX-2 inhibitors [22].

Quantitative Analysis of Events

All event monitoring studies provide a numerator (the number of reports of an event) and denominator in terms of the number of patients

Box 15.4 Points for consideration in evaluation of reported events

- The temporal relationship (time to onset)
- The clinical and pathologic characteristics of the event
- The pharmacologic plausibility based on previous knowledge of the drug and the therapeutic class, if appropriate
- Whether the event was previously reported as an adverse reaction in clinical trials or post-marketing in the UK or in other countries
- Any possible role of concomitant medications or medications taken prior to the event
- The role of the underlying or concurrent illnesses
- The effect of dechallenge or dose reduction
- The effect of rechallenge or dose increase
- Patient's characteristics, including previous medical history, such as history of drug allergies, presence of renal or hepatic impairment, etc.
- The possibility of drug interactions

and the number of patient-months of exposure to the drug; all are collected within a known time frame. This allows for event profiles over time to be examined through application of various statistical methods; such analyses are performed using "high-level" event terms from MedDRA (or historically, higher level terms from the DSRU dictionary). The trend of reports after starting treatment can be very informative: pharmacologically related side effects tend to occur early in the study (although this period may also be affected by carryover effects from previous medication), or the number of reports may rise as time passes (as with long latency adverse reactions).

Analysis by Event Counts (Incidence)

One simple but effective descriptive method is to examine the incidence of events for the whole cohort by month, by system-organ-class (SOC). Such tables can generate signals; for example, the incidence of an event in the first month or subsequent months may be unusually high (in contrast to that expected from the Summary of Product Characteristics [SPC]). An example of such a signal is gynecomastia with finasteride (Case example 15.1) [23].

Analysis by Event Rates (Incidence Densities)

Event monitoring takes advantage of the information on duration of exposure that is provided on the questionnaire and, for M-PEM, from the NHSBSA. Rates (incidence densities [IDs]) can be calculated for a given fixed time period (t) – ID_t –for all events reported in patients for a given time period and are expressed in units of first event reports per 1000 patient-months of treatment (the time between treatment start and stop dates) or observation (the time between start date and end of survey date) if pattern of drug use is continuous or intermittent, respectively.

Thus:

$$ID_t \text{ per } 1000 \text{ patient-months of treatment } (or\ observation) = \frac{N_t}{D_t} \times 1000$$

where N_t = number of 1st reports of an event during treatment (*or observation*) for period t, and

$$D_t = \frac{\text{Number of patient-days of treatment } (or\ observation) \text{ for period t}}{30}$$

where 30 defines a 30-day month.

Incident densities can be calculated for each individual month for the relevant study period, as well as combinations of months (this being dependent on the study question – see later) and all months combined (ID_A). Ideally, the exposure time would be censored at the time of the first event. However, since there are a large number of health outcomes of interest and the

Case example 15.1 Finasteride and gynecomastia

Background

- Indicated for the treatment and control of benign prostatic hyperplasia (BPH).
- Finasteride is an inhibitor of 5-alpha-reductase, which catalyzes the conversion of testosterone to dihydrotestosterone (DHT).
- Company literature at time of marketing indicated that the most frequent AEs were related to sexual function and that there were feminizing effects; gynecomastia was not included in the data sheet when the product was launched.

Question

- How did PEM help in identifying this signal?

Approach

- Quantitative methods included constructing a list of events, by system–organ class and according to treatment status, with denominators of number of male patients still in the study, by month. Data were then compared across other drugs within the PEM base.
- Qualitative methods involved further evaluation and characterization of the event.

Results

- The PEM cohort comprised 14 767 males (mean age 69 years).
- Reports of impotence/ejaculatory failure and decreased libido were received in relation to the first and all subsequent months of treatment, but reports of gynecomastia were only rarely received before the fifth month of therapy (Figure 15.5).
- To assess whether gynecomastia was an adverse event with finasteride, the data for 41 completed PEM studies were examined for reports of this event; only 17 of these 41 studies had gynecomastia. There were 42 reports (39 on drug) for finasteride (incidence rate 0.26 per 1000 patient-months of treatment) compared to 75 (56 on drug) for the other 17 studies combined (incidence range 0.03–0.23 per 1000 patient-months of treatment). These results strengthened the signal further.
- Follow-up of these cases of gynecomastia and 12 "potential" cases (with signs and symptoms of the condition based on other events) reported that the gynecomastia resolved on dechallenge in 15 of the 31 men in whom

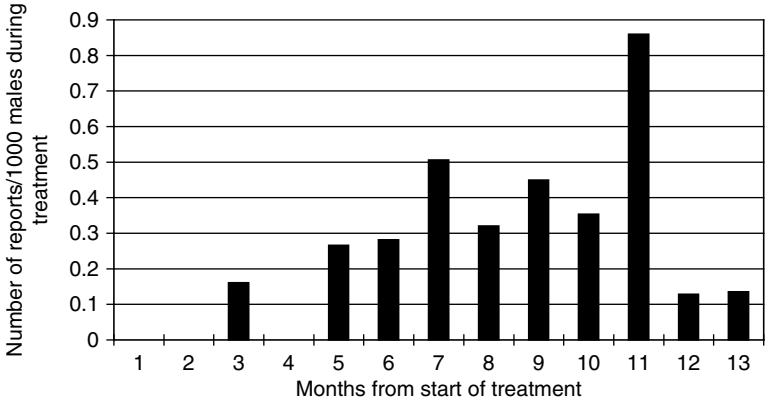


Figure 15.5 Reports of gynecomastia during treatment with finasteride.

finasteride was given in the absence of other relevant concomitant therapy.

Strengths

- This shows the complementary and essential nature of qualitative and quantitative methods in assessing risk.

Limitations

- The incidence rate calculated was a crude measure, and there was no ability to control for confounding.
- Follow-up was not systematic for all reports in the database, so only additional data were obtain for the cases exposed to finasteride.

Key points

- This postmarketing surveillance study generated a signal that was not identified in premarketing clinical trials of finasteride.
- While the incidence measure may have been subject to bias, the outcome was that the data sheet was amended.

censoring would be different for each outcome, the denominator for the crude ID does not initially account for censoring. Any subsequent analysis of particular signals would use appropriately censored denominator data. One of the many different evaluations possible is to rank these data in descending order of the estimate of ID_1 (ID in month 1) and to examine the frequency of clinical events unlikely to be confounded by indication. See Case example 15.2 which describes the PEM study of drospirenone/ethinyl estradiol [24].

Comparisons of Event Rates (Incidence Densities)

Calculating measures of impact (ID rate differences) or measures of effect (rate ratios) are other quantitative evaluations that can be used to identify events that occur significantly more frequently soon after starting the study drug. The null hypothesis is that the incidence rates are constant between the two time periods in a fixed cohort, that is, the events are not related to treatment in any way; the alternative hypothesis is that the incidence rates are different between the two time periods in a fixed cohort. In rejecting the null hypothesis where substantial differences are observed, this could be explained by a number of factors, including drug treatment. Most frequently, for each reported event, the difference or ratio between the IDs in the first month after starting treatment and the IDs for months two to six ($ID_1 - ID_{2\text{-}6}$) is calculated to allow the examination of the null hypothesis that the rate for the event is not increasing or decreasing between these two time periods.

A confidence interval (CI) (99% or 95%) is applied to the difference or ratio in the rates between months as specified above; these are computed based on the Normal approximation. Thus, where the $ID_1 - ID_{2\text{-}6}$ value for an event is positive, or $ID_1/ID_{2\text{-}6}$ is above one and the confidence limits around the point estimate exclude the null value (zero or one respectively), the rate of events in month 1 is significantly greater than the rate of events in month 2–6 combined. This result can be considered a signal for an event associated with starting treatment with the study drug.

In comparing these two time periods, the assumption is made that, given an event, its reporting is equivalent in both periods in a fixed cohort. It is recognized that there are a number of limitations to this method of examining the data – these will be discussed subsequently. Similarly, ID differences or ratios can be used to identify events that have a delayed onset, for example where the $ID_1 - ID_{2\text{-}6}$ value for an event is negative, or the $ID_1/ID_{2\text{-}6}$ is less than one and the confidence limits around the point estimate exclude the null value (0 or 1 respectively).

Case example 15.2 Yasmin®

Background
- Yasmin is a combined oral contraceptive (COC) containing ethinyl estradiol and a new progestogen, drospirenone; it was launched in the UK in May 2002.
- While the association between estrogen-containing oral contraceptives and venous thromboembolism (VTE) is well established, VTE is a rare event in young women and the risk associated with a new COC cannot generally be determined from clinical trials.
- The initial prescribing information for Yasmin stated that "It is not yet known how Yasmin influences the risk of VTE compared with other oral contraceptives."

Question
- How can prescription event monitoring help to evaluate the risk of this new drug?

Approach:
- Obtain prescriptions for all new users in England, which avoids the selection bias inherent in premarketing clinical trials.
- Obtain outcomes, which are events, regardless of causality, reported by the patients' doctors (who have access to the patients' health information in both primary care and hospital contact).
- Apply qualitative and quantitative methods to generate and test hypotheses.

Results
- The PEM study for Yasmin identified 13 cases (deep vein thrombosis 5; pulmonary embolism 8) in 15 645 females using Yasmin, with a crude incidence rate of 13.7 cases per 10 000 woman-years (95% CI 7.3–23.4).
- Each of the cases had one or more possible risk factors for VTE.

Strengths
- The PEM allowed for a rapid assessment of risk; to our knowledge, this was the first description of cases of DVT and PE in users of Yasmin in the primary care setting in England.

Limitations
- Although an incidence rate has been calculated, there was no control group and no ability to account for confounding.
- Cases all had risk factors for VTE and therefore the events may not have been related to the drug.

Key points
- While premarketing clinical trials identified many aspects of the safety of Yasmin, apparently no cases of VTE were reported.
- An association between COCs and VTE has been recognized for more than 40 years so it is important to know whether a new COC is associated with VTE and to what extent.
- The lack of selection bias and large numbers of women studied led to identification of women who developed VTE while taking Yasmin, all of whom had risk factors for VTE. The PEM study raised the need for special consideration before women with risk factors for VTE take Yasmin.
- While the incidence of VTE in the PEM study may have been subject to bias, it is the first computed incidence for this condition with Yasmin. Nonetheless, it needs to examined by other studies.

In such settings, the rate of events in months 2–6 combined is considered to be significantly greater than during month 1 and this result is considered to be a signal for a delayed-onset event. These signals then require confirmation or refutation by further study. Table 15.1 shows a summary of such data from a typical PEM study of a drug (oxcarbazepine) for which

Table 15.1 Incidence densities (ID per 1000 patient-months) for standard PEM study for oxcarbazepine ranked in order of ID_1 for all events during treatment (between date of starting and stopping) where $ID_1 \geq 3$. Events associated with starting treatment are in ***bold italic.***

Higher term	N_1	N_{2-6}	ID_1	ID_{2-6}	ID_1-ID_{2-6}	99% CI	N_A (%)	ID_A	RFS	ADR
Dose increased	***95***	***104***	***49.22***	***12.91***	***36.30***	***22.90, 49.71***	***278 (12.4)***	***15.30***	-	-
Convulsion, epilepsy	***81***	***102***	***41.96***	***12.66***	***29.30***	***16.87, 41,73***	***247(11.0)***	***13.59***	***41***	***2***
Not effective	40	133	20.72	16.51	4.21	−5.00, 13.42	270 (12.0)	14.86	262	-
Drowsiness, sedation	***38***	***44***	***19.69***	***5.46***	***14.22***	***5.73, 22.72***	***103(4.6)***	***5.67***	***57***	***15***
Nausea, vomiting	***33***	***33***	***17.10***	***4.10***	***13.00***	***5.12, 20.88***	***75(3.3)***	***4.13***	***23***	***6***
Malaise, lassitude	***31***	***37***	***16.06***	***4.59***	***11.47***	***3.79, 19.14***	***83 (3.7)***	***4.57***	***44***	***9***
Dizziness	***30***	***32***	***15.54***	***3.97***	***11.57***	***4.04, 10.10***	***74 (3.3)***	***4.07***	***31***	***4***
Dose reduced	***27***	***51***	***13.99***	***6.33***	***7.66***	***0.36, 14.95***	***131 (5.8)***	***7.21***	-	-
Rash	***22***	***23***	***11.40***	***2.86***	***8.54***	***2.10, 14.98***	***60 (2.7)***	***3.30***	***32***	***5***
Headache, migraine	***21***	***32***	***10.88***	***3.97***	***6.91***	***0.53, 13.28***	***62 (2.8)***	***3.41***	***21***	***3***
Visual defect	20	37	10.36	4.59	5.77	−0.51, 12.04	73 (3.3)	4.02	26	5
Hospital referrals no admission	16	31	8.29	3.85	4.44	−1.19, 10.07	70 (2.1)	3.85	26	-
Unspecified side effects	16	24	8.29	2.98	5.31	−0.25, 10.87	56 (2.5)	3.08	51	56
Electrolyte abnormal	15	29	7.77	3.60	4.17	−1.28, 9.62	67 (3.0)	3.69	30	1
Nonsurgical admissions	14	22	7.25	2.73	4.52	−0.69, 9.73	56 (2.5)	3.08	9	-
Ataxia	10	9	5.18	1.12	4.06	−0.26, 8.39	20 (0.8)	1.10	7	3
Condition improved	8	13	4.14	1.61	2.53	−1.41, 6.48	31 (1.4)	1.71	20	-
Intolerance	8	5	4.14	0.62	3.52	−0.32, 7.38	18 (0.8)	0.99	17	3
Unsteadiness	8	14	4.14	1.74	2.41	−1.55, 6.36	29 (1.4)	1.60	6	3
Confusion	7	18	3.63	2.23	1.39	−2.39, 5.17	37 (1.7)	2.04	15	5
Fall	7	10	3.63	1.24	2.38	−1.29, 6.06	28 (1.3)	1.54	2	1
Depression	6	25	3.11	3.10	0.00	−3.63, 3.64	42 (2.1)	2.31	7	4

(*Continued*)

Table 15.1 (Continued)

Higher term	N_1	$N_{2–6}$	ID_1	$ID_{2–6}$	ID_1-$ID_{2–6}$	99% CI	N_A (%)	ID_A	RFS	ADR
Patient request	6	8	3.11	0.99	2.12	–1.28, 5.51	22 (1.0)	1.21	22	-
Pruritus	6	10	3.11	1.24	1.87	–1.55, 5.29	20 (0.8)	1.10	8	1
Tremor	6	2	3.11	0.25	2.86	–0.44, 6.16	11 (0.5)	0.61	5	2

N_1 = Total number of reports of each event during the first month of treatment.
$N_{2–6}$ = Total number of reports of each event during treatment in months 2–6.
ID_1 = Incidence density for each event during the first month of treatment (where D_1 = 1930).
$ID_{2–6}$ = Incidence density for each event during treatment months 2–6 (where $D_{2–6}$ = 8054).
ID_1–$ID_{2–6}$ = Arithmetic difference between ID_1 and $ID_{2–6}$.
99% CI = 99% confidence intervals for ID_1–$ID_{2–6}$.
N_A (%) = Total number of reports of each event (% incidence in total cohort) during total treatment period.
ID_A = Incidence density for each event for the total treatment period (where D_A = 18170).
RFS = Reason for stopping oxcarbazepine (total no. reports = 932 in 698 patients (31.1% of cohort).
ADR = Adverse drug reaction (total no. reports = 158 in 105 patients – 4.7% of cohort).

pattern of use is considered continuous; it is restricted to events reported during treatment (between start date and stop date) and with corresponding denominator of patient-months of treatment. Events associated with starting treatment are in bold.

For drugs where pattern of use is intermittent and/or short term, such summaries are also produced, but there are several differences. First, the numerator is based on total incident counts irrespective of treatment status (whether recorded during or after treatment or whether "unknown") and the denominator takes into account the observation period (between start date and end of survey date). Second, the comparator (reference) period may be restricted. Table 15.2 shows a summary of such data from a PEM study of a drug (levocetirizine) intended for short-term (<30 days) intermittent use, where the second month was considered most appropriate as the reference period.

Time to Onset

It is acknowledged that the generalized approach to segregation of time periods may not be appropriate for all events with respect to their most relevant time periods of excess. It is possible to explore the time taken for an event of interest to occur by using time-to-event analysis, thus providing an additional tool for signal generation purposes. One example is the incidence rate of venous thromboembolism (VTE) as reported in the PEM study of strontium [25].

Plotting a hazard function in a fixed cohort for events of interest is another useful method to determine whether the hazard (instantaneous risk) of the event increases or decreases with time. A constant hazard over time may be consistent with a background (i.e., not caused by the drug) event rate, whereas a nonconstant hazard over time may be an indicator of a drug–event relationship. Since it is desirable to understand the shape of the underlying survival function, parametric time-to-event models can estimate the baseline hazard function and the instantaneous change in hazard over time, and the goodness-of-fit can be assessed. An example is the examination of hazard rates of neuropsychiatric events as reported in the M-PEM study of varenicline [26].

Reasons for Stopping the Drug

All event monitoring questionnaires ask the prescriber to record the reason why the drug was stopped, where treatment cessation occurred. This is informative because it includes possible adverse reactions which the physician and/or the patient considered serious or sufficiently troublesome to stop the medication. Clinical and nonclinical reasons for stopping a drug are presented in two ways: by system-organ class by month, and also ranked by total count. The ranked reasons for discontinuation can be compared with the ranked incidence density estimates and this comparison can also generate signals. There is usually a good correlation in terms of the most frequently reported events (see examples in Tables 15.1 and 15.2).

Outcomes of Pregnancy

All pregnancies reported during event monitoring studies are followed up in order to determine the outcome in those babies exposed *in utero* to the drugs being monitored. There is interest in determining the proportion and nature of congenital anomalies in babies born to women exposed to newly marketed drugs during pregnancy, in particular in the first trimester. PEM studies have shown that from 831 such pregnancies, 557 infants were born, of whom 14 (2.5%) had congenital anomalies [27]. It is important that studying pregnancy outcomes continues in order to exclude, to the greatest extent possible, teratogenic effects of medicines (see Chapter 22).

Drug Utilization

In recent years, the focus on drug utilization has become more important in terms of safety, especially in relation to use in special populations

Table 15.2 Incidence densities (ID per 1000 patient-months) for standard PEM study for levocetirizine ranked in order of ID_1 for all events during observation (between date of starting and end of observation) in the first two months after starting treatment (where $IDobs_{m1}$ >1).

Higher term[a]	N_1	N_2	$IDobs_{m1}$	$IDobs_{m2}$	$IDobs_{m1/m2}$ (95% CI)	ADR	Reason for stopping[b]
Condition improved	1470	434	118.90	35.15	3.38 (3.04, 3.77)	NA	**1896**
No further request	640	59	51.77	4.78	10.83 (8.29, 14.40)	NA	699
Not effective	460	133	37.21	10.77	3.45 (2.84, 4.22)	NA	588
Course completed	160	29	12.94	2.35	5.51 (3.69, 8.49)	NA	189
Other drug substituted	62	26	5.02	2.11	2.38 (1.48, 3.92)	NA	88
Upper respiratory tract infection	**56**	**25**	**4.53**	**2.03**	**2.24 (1.37, 3.74)**	**0**	**4**
Drowsiness, sedation	**46**	**4**	**3.72**	**0.32**	**11.48 (4.19, 43.93)**	**5**	**43**
Headache, migraine	**22**	**9**	**1.78**	**0.73**	**2.44 (1.08, 6.02)**	**2**	**6**
Hospital referrals no admission	22	11	1.78	0.89	2.00 (0.93, 4.56)	**0**	**10**
Noncompliance	21	2	1.70	0.16	10.49 (2.56, 92.24)	NA	18
Rash	20	9	1.62	0.73	2.22 (0.97, 5.53)	1	8
Pregnancy	**11**	**2**	**1.53**	**0.28**	**5.49 (1.20, 51.00)**	NA	**5**
Urinary tract infection	18	12	1.46	0.97	1.50 (0.68, 3.41)	0	0
Anxiety	**17**	**4**	**1.38**	**0.32**	**4.24 (1.38, 17.34)**	**0**	**0**
Lower respiratory tract infection	17	21	1.38	1.70	0.81 (0.40, 1.61)	0	0
Pain joint	15	8	1.21	0.65	1.87 (0.75, 5.10)	0	0
Nonformulary product	13	8	1.05	0.65	1.62 (0.62, 4.52)	NA	21

[a] Clinical events associated with starting treatment are highlighted in bold.
[b] 3732 reasons for stopping during months 1 and 2 of observation, of 5509 (for whole study period).
ADR, events recorded as adverse drug reactions (25 reports [in months 1 and 2 of observation] of 31 [for whole study observation period]); ID, incidence densities; $IDobs_{m1}$, incidence density for each event during observation month 1; $IDobs_{m2}$, incidence density for each event during observation month 2; $IDobs_{m1/m2}$, relative difference between $IDobs_1$ and $IDobs_2$; N_1, total number of first reports of each event during observation in month 1; N_2, total number of first reports of each event during observation in month 2; NA, not applicable.

and off-label use. There is interest in determining the prevalence of use in such populations, in addition to adverse event profiles. For example, an M-PEM study on extended-release quetiapine found that the prevalence of off-label prescribing in terms of indication and high doses was common, as was use in special populations such as the very elderly [28].

Strengths

Representativeness and Size

Event monitoring uses a noninterventional cohort design; it does not interfere with the decision to prescribe a medication and information is collected *after* the prescribing decision has been made and implemented. This means that in event monitoring, data are collected on patients who have received the study drug because the doctor considered it the most appropriate treatment for that patient, as in everyday real-world clinical practice the patient would have been prescribed the drug regardless of whether they were included in the event monitoring study. In contrast to clinical trials, there are no predefined selection criteria in terms of prescribers or patients based on specific characteristics. These study characteristics make it likely that the event monitoring cohort is representative of all patients who have started the study drug under similar circumstances on a national scale in England. As a result, generalizability is ensured, unlike in many clinical trials.

In terms of size, event monitoring has been shown to be successful in collecting information on very large cohorts of patients exposed to new treatments in the UK. Cohorts of 10 000 patients were regularly established during original PEM studies and current M-PEM and SCEM studies have gathered appropriate sample sizes based on the objectives of the study.

Variable Direction of Investigation

Event monitoring consists of both prospective data collection and retrospective collection where required. Patient medical history can be studied, in addition to prospective data collection on specific events or conditions of interest.

Exposure Data

It is important for drug safety that exposure windows are appropriately calculated to minimize biased estimates of association (ID differences) or estimates of effects (ID ratios) where internal *a priori* comparisons are undertaken. In M-PEM, exposure data are derived from dispensed prescriptions, with validation from prescribers through confirmation of such data on the questionnaires. Considering the large proportion of patients who are prescribed a medication but do not get the prescription filled [29], this is an advantage in that M-PEM exposure data are more accurate than those derived from records of physician-issued prescriptions (which are not always dispensed), as held in some pharmacoepidemiologic databases. In SCEM studies, exposure data are not based on dispensed prescriptions; however, since patients are under specialist care in such studies, they are likely to be more closely monitored in terms of medication use and adherence. Many patients will also visit their specialist at frequent intervals, unlike in the primary care setting.

Outcome Data

In event monitoring, the study design collects information on events regardless of causality. Therefore, event monitoring is able to identify signals of adverse reactions or syndromes which none of the participating physicians suspect to have been due to an ADR [30]. If the physician suspects an event is due to the use of a drug, then they can specify this on the questionnaire. In addition, the nonpassive design

prompts the physician to fill in the questionnaire regardless of any events experienced and does not rely on them taking the initiative to report. This "prompting" effect of event monitoring is critical because ADR reporting is enhanced in event monitoring compared to the passive Yellow Card spontaneous ADR reporting system in the UK [31].

Signal Strengthening

Event monitoring can be used to identify patients with potential ADRs who can be studied further. Case series can be used to examine clinical characteristics of particular adverse events within patients at aggregate level. In addition, comparisons within the DSRU database are also possible and can be conducted to refine signals [32,33]. These comparisons are appropriate because the database is composed of new drug user populations with exposure in the immediate postmarketing period since introduction of each product. It is also possible to conduct external comparisons using demographic data of the population as a whole, such as standardized mortality ratio [34]. However, in contrast to event monitoring data, some medical record databases (such as the Clinical Practice Research Datalink or The Health Improvement Network, see Chapter 13) have limited data on recently introduced products, which precludes reliable comparisons being made because of small sizes of the population exposed.

Definitive Answers

Event monitoring can also confirm or refute drug safety signals. Various approaches can be used, such as nesting a case–control study within an event monitoring cohort. This method overcomes some of the disadvantages associated with nonnested case–control studies while incorporating some of the advantages of a cohort study [35]. As a pharmacoepidemiologic tool for risk management plans, the design potentially offers impressive reductions in costs and efforts of data collection and analysis compared with the full cohort approach, with relatively minor loss in statistical efficiency. Event monitoring cohorts provide opportunities to conduct such nested case–control studies, for example, for patients who develop selected ADRs and matched patients who receive the same drug without developing ADRs. One example is the possible association between the use of an atypical antipsychotic and extrapyramidal symptoms. Other approaches include comparisons between event monitoring cohorts, such as comparing the risk of drowsiness and sedation between two antihistamines [32].

Use of Reference (Contextual) Cohorts

Event monitoring studies can use reference (contextual) cohorts when it is considered inappropriate to use counterfactual comparative cohorts. The objective of such contextual cohorts is to examine differences in characteristics between patients who receive a new medicine and those who are on standard therapy. This is undertaken when it is expected that there are substantial differences in the characteristics of the two cohorts which will impact on comparative safety and where this cannot be adequately addressed by methods usually used in pharmacoepidemiology to handle such imbalances [36].

Participation in Research

For physicians in the UK, research and academic medical practice are considered noncore activities and therefore receive no payment from the NHS. Although GPs have a duty of care to report ADRs and cooperate with requests for information from organizations monitoring public health, those GPs who participate in event monitoring studies do so on behalf of research and not for monetary interest – the remuneration received for completion of forms barely covers administration costs for

M-PEM, while for SCEM studies remuneration is paid to the NHS trust and not to the physician. This is also the case for other pharmacovigilance activities such as the Yellow Card spontaneous reporting scheme, where no money is received for participation.

Since the early 1980s, close contact has been fostered between the research staff in the DSRU and the reporting doctors; focus groups and advisory committees are used to ensure maximum response and maintain good relationships with physicians. This facilitates the gathering of supplementary information on important events, pregnancies, deaths, etc. (see Table 15.1 and Box 15.1), which allows for the maximal clinical understanding of biases, the natural history of ADRs, and other important risk factors (which could be potential confounders for prespecified internal comparisons).

Limitations

Single-Group Cohort Design

As highlighted earlier, event monitoring uses a simple single-group cohort design where subjects have been assembled based on a common exposure (the particular medication under surveillance). Compared to the "classic" cohort design with an explicit comparator, it is more efficient in terms of resources. In event monitoring studies with absence of data on an unexposed comparator, calculating measures of effect (relative risks) is restricted to internal comparisons between subgroups defined by particular characteristics, or external comparisons to carefully selected data sources. Moreover, as stated above, in recent years, the DSRU has modified the event monitoring design to include a contextual comparator within certain studies; for example, a warfarin cohort was also recruited for the Rivaroxaban Observational Safety Evaluation (ROSE) SCEM study. This comparator was chosen to inform on the adoption of rivaroxaban into clinical practice and variation in determinants of treatment choices though it was not used to calculate differences in risk between treatment groups.

Bias

Validity in observational studies is an important consideration and selection bias can be introduced if a nonrepresentative sample of the population is recruited [37]. Such error cannot be adjusted for. How an event monitoring cohort differs to all other patients with the same indication receiving other healthcare in the UK cannot be assessed since, as mentioned previously, event monitoring does not always monitor an unexposed cohort concurrently. Channeling of new drugs can also introduce selection bias through preferential prescribing. Patterns of adoption of a new drug cannot be predicted, and while it may be examined in event monitoring, it cannot be controlled.

Nonresponse bias is another form of selection bias which is possible since not all M-PEM questionnaires are returned. The response rates for M-PEM studies are usually over 50%, which is comparable to response rates reported elsewhere for GP postal surveys [38] and higher than the reporting rates of suspected ADRs in the Yellow Card scheme [31,39]. In M-PEM, it is not known whether the prescribers who do not respond (and their patients) differ from those prescribers (and their patients) who do participate.

Underreporting, including underreporting of serious and fatal adverse events, is possible in event monitoring since it depends on reporting by physicians. Information bias in terms of misclassification of outcome and exposure is also possible since the data depend on the accuracy and thoroughness of the physicians in diagnosis, record keeping, and reporting.

In event monitoring, exposure misclassification may be introduced through inaccurate calculation of exposure (time on treatment).

It is important because inappropriately calculated exposure windows can result in a biased estimate of effect, particularly if unnecessarily long because relative differences get diluted as the time window widens and a potential signal may be lost. In M-PEM, exposure is calculated from dispensed prescriptions, which means that exposure data used for M-PEM are more accurate than exposure data based on physician-issued prescriptions alone. In SCEM, exposure data are based on the prescription date from specialist care, which is likely to be closely monitored by the specialist and in some cases event administered within a hospital setting. Nevertheless, patients may not take all the dispensed medication. In this regard, the misclassification of exposure is likely to be non-differential, being the same across the new drug cohorts, and the effect estimate (ID rate difference/ratio) biased towards the null. As for observational studies in general, assumptions are made regarding compliance and for drugs used for chronic conditions, the assumption is made that individual patients take the medication up to the end of treatment (or stop date) unless otherwise indicated. However, for drugs taken over long periods of time, repeat dispensing is not an absolute proof of good compliance but is a usually a good marker of compliance.

Confounding

Event monitoring is affected by a limitation common to all observational studies – the inability to control for factors that might differ between groups being compared [37]. Hypotheses generated by event monitoring may be further explored using traditional hypothesis-testing techniques, such as case–control methods. However, it is important to acknowledge that the range of data that can be collected on important co-variates for *all* possible outcomes may be limited. In such cases, when examining relationships between exposure and outcomes within a case series, or when conducting comparisons between subgroup populations within a drug or between drugs to strengthen or refine signals, data may be incomplete or missing and residual confounding is likely. Nevertheless, M-PEM and SCEM provide considerable opportunities to enhance collection of supplementary data on important risk factors.

Limited Statistical Power and Sample Size

It is possible to calculate power and sample size for a single cohort study, provided one has a hypothesis about the effect size and the background rate involved. However, the detection of rare ADRs is not always possible, even with cohorts of 10 000–15 000 patients. Due consideration should also be given to the nature of event monitoring statistical analysis which involves running routine multiple comparisons whereby event ID difference or ratio statistics are generated to examine the null hypothesis that event rates are constant between two time periods. A minimum 95% confidence interval has routinely been used in event monitoring to aid decision making, but in terms of signal generation for safety surveillance, there is still the chance of a type II error, that is, of missing a difference that really is there.

Particular Applications

Signal Strengthening

Signal Strengthening through Quantitative Evaluation

Once a signal has been recognized, supplementary analysis is required to further characterize important attributes. As highlighted previously, PEM provides the opportunity for further collection of detailed information on reported events and allows systematic review of individual case reports and aggregate data.

One important example of follow-up exploration in relation to a long-latency adverse event concerned visual field defects in patients receiving long-term treatment with vigabatrin [21]. The initial PEM study showed three cases of bilateral, irreversible peripheral field defects, whereas no similar reports occurred with other antiepileptic drugs or in any of the other drugs already monitored by PEM. A follow-up exploration with a repeat questionnaire, sent to the doctors whose patients had received vigabatrin for over six months, showed that the incidence of this serious event was much higher and that many of the relevant patients had objective evidence of visual field defects. Another example of signal follow-up is given in Case study 15.1 in relation to gynecomastia and finasteride [23].

Signal Strengthening and Hypothesis Testing

Comparison of Event Rates and Risks

Comparisons can be used to give estimates of relative measures of associations (e.g., relative risk) and are often associated with hypothesis testing. However, such pharmacoepidemiologic methods can be used to explore or strengthen signals as an extension of postmarketing safety surveillance. In event monitoring, a variety of targeted comparisons of event rates and risks occurring between different patient populations are conducted to explore apparent associations. These can be segregated into two sorts: using internal comparators such as subsets of patients within the same drug cohort or between drugs within the same therapeutic class; or using an external comparator. The research question being asked (usually) determines which pharmacoepidemiologic design for these comparisons should be used and the most appropriate statistical analyses required.

Various methods are applied to enable nested internal comparisons between subgroups defined by particular characteristics. Such comparisons can be conducted using event monitoring data, including simple stratification, "before and after" matched analyses, multivariate modeling, and standardization.

Simple stratification

Through simple stratification, event profiles in subgroups of patients can be examined, and rates of preselected events compared between these subgroups by calculating crude relative risks or rate ratios. The assumption is that all other characteristics are constant because the subgroups are nested within the new user cohort, although residual confounding is likely (as discussed above).

One example, for which the aim was to look for evidence of channeling a new drug to problem patients, was to examine and compare the frequency of gastrointestinal events reported in a PEM study of the COX-2 selective inhibitor celecoxib in those patients with gastrointestinal (GI) risk factors (past history of GI conditions, gastroirritant drugs, use of concomitant gastroprotective agents) to those without [40]. In this example the null hypothesis was that risk was the same in both subgroups. In this study, significantly higher rates of GI events were observed in patients with risk factors, which supports the possibility of channeling bias.

In another example, a M-PEM study examined the frequency of major and minor depressive episodes with rimonabant (an obesity treatment) after starting treatment. This was a safety concern identified at the time of marketing approval and so patients with a previous history of psychiatric illness were examined separately to those without a previous history. In patients without a previous history of psychiatric illness, there were more major and minor depressive episodes in the six months after starting treatment compared with the six months before starting treatment with rimonabant (relative risk [RR] 1.7; 95% CI 1.2–2.3 and RR 1.33; 95% CI 1.20–1.48, respectively). The same was not found for those who had a previous history of psychiatric illness [41]. Marketing authorization for rimonabant was withdrawn in

October 2008, mainly because the psychiatric adverse effects could not be addressed by further risk minimization [41].

Before and After Studies

"Before and after" studies compare the rate of particular outcomes during a defined period of exposure (or observation) after starting the study drug with those rates in the same individuals during a defined period of observation before starting, using a matched pair analysis. The null hypothesis is that event rates are the same before and after starting treatment.

One example was the examination of rates of respiratory events with the introduction of a new chlorofluorocarbon (CFC)-free formulation of an anticholinergic (ipratropium) metered-dose inhaler (MDI) in populations who were "switchers" from the original MDI and those naïve to ipratropium treatment [42]. The analyses suggested that characteristics of these two subpopulations differed such that naïve patients were more likely to be children, have an indication of asthma, and have milder disease severity, while switchers were more likely to be adults, have an indication of COPD, and have more severe disease. Such differences have an important impact on ongoing evaluation of benefit/risk balance of the new formulation. Common respiratory events occurred at higher rates after starting treatment than before for switchers, for example lower respiratory tract infection (LRTI) (RR 1.45; 99% CI 1.17–1.81) and worsening asthma (RR 1.58; 99% CI 1.00–2.51). Of these events, only LRTI was significant for naïve patients (RR 1.42; 99% CI 1.04–1.95).

Modeling

Multivariable modeling examines the potential effect of one variable on the outcome of interest while controlling for many other variables. An example of multivariable conditional logistic regression modeling was a within-PEM study comparison to examine the risk of pioglitazone treatment combinations (with insulin or other antidiabetic agents) on risk of hypoglycemia [43]. The null hypothesis was that the risk of this outcome was the same regardless of treatment. Pioglitazone may be used alone or in combination with a sulfonylurea, metformin, or insulin as an adjunct to diet and exercise for the management of type 2 (noninsulin-dependent) diabetes mellitus (NIDDM). The summary of product characteristics states that hypoglycemia is common when pioglitazone is administrated in combination with insulin and very common during the triple combination treatment with metformin and sulfonylurea [43]. Patients taking combination therapy with sulfonylurea or insulin were estimated to have approximately three and four times the hazard of having an event compared with those who were not taking these adjunctive therapies (hazard ratio [HR] 3.11; 95% CI 1.64–5.88; HR 4.15; 95% CI 1.74–9.91, respectively). Patients treated with adjunctive metformin were 25% less likely to experience hypoglycemia than those who did not take concomitant metformin (HR 0.75; 95% CI 0.44–1.27). This suggests that patients taking pioglitazone with insulin or sulfonylurea had higher risks than those on pioglitazone monotherapy [43].

An example of the application of Poisson regression modeling (which takes different exposure durations into account) was to examine whether there was a difference in incidence rates for thromboembolic (TE; cardiovascular, cerebrovascular, and peripheral venous) events reported for patients dispensed rofecoxib and meloxicam, because of the unexpected association shown in a clinical trial [44]. The null hypothesis was that event rates were the same regardless of drug. This study reported a relative increase in the rate of cerebrovascular TE events (RR 1.68; 95%CI 1.15–2.46) and a relative reduction in peripheral venous thrombotic events (RR 0.29; 95% CI 0.11–0.78) for rofecoxib compared to meloxicam, after adjusting for age and sex. There was no difference in the rate of

cardiovascular thrombotic events. This particular example shows how the DSRU event monitoring database provides a resource to evaluate signals and hypotheses generated by other sources. Another example is the comparison of mortality and rates of cardiac arrhythmias with atypical antipsychotic drugs [45].

Standardization

Where appropriate, comparisons are made between patients identified within an event monitoring study and an external reference group, if a suitable internal reference cohort cannot be found and the research question requires the result to be contextual. For instance, calculation of standardized mortality ratio (SMR) is an indirect method of adjusting a mortality rate to that the observed death rate can be compared to that expected if the study cohort has the same characteristics of the reference cohort. Thus, the SMR is the ratio of observed deaths to expected deaths. Hence through using this indirect method of standardization, the expected deaths in an event monitoring cohort can be calculated using information on the general population age-specific rates.

Following concerns about cardiovascular safety with sildenafil, the mortality from ischemic heart disease in users of sildenafil in a PEM study was compared with external epidemiologic data for men in England [34]. The SMR for deaths reported to have been caused by ischemic heart disease (IHD) in the sildenafil PEM cohort was 31.41 (95% CI 18.29–50.29, based on Poisson error factors), indicating that the point estimate mortality in the cohort was 68.6% lower than that for males in England in 1998. However, the 95% CI is wide and there is no evidence to suggest a higher incidence of fatal IHD among men in England taking sildenafil. Similarly, death from ischemic heart disease in the bupropion PEM (when used for smoking cessation) was compared with external data and showed no difference in the SMR [46]. Obviously, there is higher potential for bias when using external comparators than comparisons undertaken between event monitoring studies, principally due to differences in study design and data collection methods; results of external comparisons must therefore be interpreted very carefully.

Automated Signal Generation

The DSRU has previously explored the use of data-mining disproportionality methods that are commonly used in pharmacovigilance (see Chapters 10 and 27) as a possible additional quantitative tool in event monitoring for signal generation, because of the large number of drug–event combinations held in the DSRU database. Feasibility studies have employed proportional reporting ratios (PRRs) [47] to quantify the ratio of observed-to-expected PEM event reports to explore historical signals – for example, Stevens–Johnson syndrome with the antiepileptic drug lamotrigine [48,49]. An extension to this method which integrates available PEM data on exposure to calculate the incidence rate ratio (IRR) has also been examined and applied to investigating new signals such as exacerbation of colitis with rofecoxib [50].

There are a number of methodologic issues which may influence whether a signal is generated; these include selection of comparator(s), signal threshold, variation in duration of study observation period, handling small event counts, and the level of dictionary terms used, such as higher- or lower-level terms. However, with refinement, automated signal generation can be a useful tool to support signal generation for event monitoring through other quantitative methods as described above.

Drug Utilization

Drug utilization research (see Chapter 18) is an essential part of pharmacoepidemiology, as it describes the extent, nature, and determinants of drug exposure at the patient level.

Data from event monitoring studies can inform about prescriber adoption of new drugs. The demographic and clinical characteristics of new users can be described and examined in relation to signals of off-label use, for example, indications, dose, and conditions or other factors that are contraindicated or special warnings for use.

An example is the PEM study of testosterone patch indicated for hypoactive sexual desire disorder in surgically menopausal women receiving concomitant estrogen therapy. Given the narrow prescribing indication, only 20.9% of the cohort were being prescribed the patch according to the manufacturer's recommendations [51].

In addition, PEM studies can examine aspects of adherence to prescribing guidelines. For example, in both PEM studies of rofecoxib and celecoxib, not only were high proportions of new users recorded as NSAID naïve (approximately 50%), but also a significant proportion (38% and 46%, respectively) had no prior history of gastrointestinal conditions (i.e., were at low risk) [40,52]. These observations were discordant with national NSAID prescribing practice during the time these drugs were first marketed, and agree with findings from elsewhere [53]. In addition, in the M-PEM study of extended-release quetiapine, off-label prescribing in terms of indication and high doses was common, as was use in special populations such as the very elderly [28].

Monitoring Drug Safety in Children

The safety of medication use in children is of major public and regulatory interest. However, there is a significant lack of safety data when a new drug is launched because of the limited number of clinical trials in this population. Furthermore, postmarketing pharmacovigilance systems for this population face significant challenges, particularly in regard to data capture of "off-label" use. European regulations have been issued that oblige pharmaceutical companies to submit a Pediatric Investigation Plan (PIP) for all new compounds, indications, and formulations. Pediatric pharmacovigilance activities have to be included in the benefit/risk management plan (see Chapter 24) and other pharmacovigilance activities. Therefore, pharmacovigilance tools may need to be adapted to examine specific issues associated with this special population.

Given that event monitoring studies capture drug usage under "real-life" conditions in general practice, including off-label prescribing to the pediatric population, it is possible to explore differences in risk profiles between children and adults using these studies. An example is a study which compared the adverse event profiles of children and adults taking lamotrigine, using modified signal detection methods [54]. Data were stratified by age and IDs were examined between two time periods after starting treatment (month 1 and months 2–6 combined). Proportional reporting ratios (PRR) and incidence rate ratios compared the risk of adverse events between adults (n=7379) and children (n=2457). Rash (PRR 1.2) and Stevens–Johnson syndrome (PRR 4.5) were more commonly reported in children, and confusion more frequently in adults (PRR 6.3). In children, 33% of events suspected to be ADRs (15/46) were reported to the regulatory authority compared with 44% (56/128 reports) in adults. In another PEM study of the oral iron chelating agent deferasirox, reported events were stratified by pediatric age groups relevant to the licensed indications for use, allowing examination of the safety profile within these specific groups [55].

Quantifying ADR Reporting

The characteristics of ADR reporting have been examined previously within the DSRU database. Two studies conducted on different PEM studies and over different periods of time compared events that were considered as ADRs by doctors

reported in PEM, with spontaneous reports sent by the same doctors to the regulatory authority [31,39]. The first study showed that 275 of 3045 suspected ADRs reported on the questionnaires of 10 PEM studies (9%) (95% CI 8.00–10.00) were spontaneously reported to the UK regulatory authority [39]. The estimate was similar in the second study conducted in 2001, based on 15 other completed PEM studies. In that study, 376 of 4211 ADRs (9%) (95% CI 8–9.8) reported on the PEM questionnaires were reported on Yellow Cards to the CSM [49]. This represents an underreporting rate of 91% in both studies. It is of interest that a higher proportion of serious than nonserious reactions were reported to the regulatory authority by doctors in both studies (53.0% vs 8.4% and 22.8% vs 8.3%, respectively), which suggests that doctors use the spontaneous adverse reaction reporting system more energetically when reporting those serious reactions that concern them most.

It is possible to use event monitoring to study general patterns of ADRs. Our studies in this area have also shown that, in general practice in England, suspected ADRs to newly marketed drugs are recorded more often in adults aged between 30 and 59 years and are 60% more common in women than in men [56]. Possible explanations for these observations include increased frequency of consulting rates for women compared to men, pharmacologic differences between men and women in distribution of medication in the body, and increased rates of recording of clinical events with age. Another important factor is prescriber type – whether they routinely participate in postmarketing studies or not.

Supporting Pharmacovigilance Risk Management Plans

The management of risk of medicines requires identification, measurement, and assessment of risk, followed by benefit/risk evaluation, then taking actions to eliminate or reduce the risk, followed by methods to monitor that the actions taken achieve their objectives. Event monitoring not only contributes to the identification and measurement of risks of medicines but, with some additions, can examine how the risks of medicines are being managed in real-world clinical settings.

An example of such a study was conducted to monitor the introduction of carvedilol for the treatment of cardiac failure [57]. The product (a combined alpha- and beta-adrenergic blocker) has been used for the treatment of angina and hypertension for some time, but there was concern about its appropriate use for cardiac failure in the community. The aim of the M-PEM study was to monitor how the product was being managed in the community; for example, what clinical investigations were undertaken prior to starting the drug, who supervised the dose titration (GP or specialist), was the drug given to patients with the appropriate severity of heart failure, etc. The design included sending an eligibility questionnaire followed by up to three detailed questionnaires for a period of up to two years. Overall, regulatory guidelines for the use and risk management of carvedilol were mostly adhered to [55].

Since risk management of medicines became a regulatory requirement in Europe in 2005, a number of event monitoring studies have been undertaken to address specific questions related to detailed examination of particular adverse events and studying drug utilization patterns (Table 15.3). Such studies support the construction of risk management plans by providing opportunities for a number of additional research applications which can be used to generate signals of potential ADRs and to further evaluate safety concerns identified by other pharmacovigilance methods or arising from regulatory concerns. Their customized sample size is advantageous in terms of study conduct, limiting costs and providing timely information to the dynamic risk management process. Thus, they should be considered a valuable tool when

Table 15.3 Design and applications of event monitoring methods.

Type	Method	Applications	Examples of completed or ongoing M-PEM studies
Special populations	Patients identified according to prespecified criteria (age, sex, indication) through use of eligibility questionnaire	• New indications • License extensions • Reclassifications • New formulations • Switching • Regulatory intervention • Hospital initiations • Health outcomes management	Carvedilol [57] – Licensed for angina and hypertension with extension in 1998 to treat mild to moderate chronic heart failure. M-PEM monitored compliance with UK regulatory authority's request for monitoring the use and safety of carvedilol in heart failure in clinical practice, and assessed clinical risk management of patients within the new indication Travoprost – Eye drops initially approved for second-line use in the treatment of ocular hypertension in open-angle glaucoma; license extension to first-line use was granted in 2003. M-PEM monitored long-term development of discoloration of the iris CFC-free MDIs (Flixotide Evohaler® and Seretide Evohaler®) [60,61] – EMA produced guidelines for the conduct of postmarketing surveillance studies to assess the introduction of CFC-free inhalers. M-PEMs collected data 3 months before and after exposure to allow event comparison before and after starting
Drug utilization	Collection of data on extent, nature, and determinants of drug use and prescribing over time	• Adherence to prescribing recommendations/ clinical guidelines • Exploration of off-label use • Characterization of real-life populations	Atomoxetine [62] –Licensed in the UK in May 2004 for the treatment of attention-deficit/hyperactivity disorder in children (6+ years) and adolescents. M-PEM monitored drug utilization with targeted data capture on psychiatric events, convulsions, abnormal liver function, and selected cardiovascular events Ivabradine – Licensed in the UK in 2006; indicated for the treatment of chronic stable angina pectoris in patients with normal sinus rhythm, with a contraindication or intolerance for beta-blockers. M-PEM investigated its use in relation to diseases/conditions that are contraindicated or warnings for use
Targeted event surveillance	Collection of data on relevant risk factors	• Hypothesis strengthening/testing	Quetiapine extended-release formulation [28] – First marketed in the UK in September 2008; indicated for the treatment of schizophrenia and manic episodes associated with bipolar disorder with license extension for add-on therapy for major depressive disorder. M-PEM included prospective nested matched case–control study to explore relationship between dose and events of somnolence and EPS. SCEM study examined safety and use in the specialist care setting Rivaroxaban – First marketed in the UK in December 2011 for the new indication of prevention of stroke and systemic embolism in adult patients with nonvalvular atrial fibrillation and the treatment of deep vein thrombosis (DVT) and pulmonary embolism (PE), and prevention of recurrent DVT and PE in adults. Complementary M-PEM and SCEM studies conducted to examine safety in both the primary care setting and the specialist care setting

CFC, chlorofluorocarbon; MDI, metered-dose inhaler.

developing a risk management plan for the evaluation of the safety of a new medicine.

The DSRU is registered within the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Database of Research resources, which serves as a central resource for both researchers and study sponsors seeking to identify organizations and datasets for conducting specific pharmacoepidemiologic and pharmacovigilance studies in Europe. The protocols for many of the individual studies conducted by the DSRU are also registered with the ENCePP.

The Future

In the future, event monitoring aims to utilize improvements in information technology, by collecting data electronically through tools such as online surveys. Additionally, plans are under way to complement the information obtained through event monitoring studies with data from across Europe, through collaborative networks spanning multiple countries.

Electronic Data Capture

Historically, all event monitoring studies at the DSRU have captured data through paper surveys; however, advances in information technology now allow data capture through a variety of methods, such as online surveys, text messaging, mobile applications, and email.

Several recent studies at the DSRU have tested these methods; a pilot study on the safety of the intranasal quadrivalent live attenuated influenza vaccine (QLAIV) was conducted. The primary objective of the study was to estimate the crude incidence rate of adverse events of interest (AEIs) following vaccination with the nasal QLAIV early in the 2014–2015 influenza season in children and adolescents in England. Participant outcomes, validated by a healthcare professional (general practitioner) where appropriate, were captured through questionnaires sent by the participant's method of choice: surface mail, telephone or web survey with email reminders. In total, 53.8% of those who consented to participate chose to complete questionnaires online [58]. The results of this pilot study indicate that offering electronic data capture is important in new studies and may help to improve response rate.

Another study examined the feasibility of rapidly monitoring safety with the new swine flu vaccines in the UK. Email and text messaging were two forms of data capture offered in this study. The methodology and use of modern technologies to collect safety data from large numbers of patients were considered successful and indicated that these could be used again for future studies [59] The questionnaire format of event monitoring studies is highly compatible with electronic data capture and the success of previous studies utilizing these techniques indicates that these are appropriate for the future of event monitoring.

Multicountry Network Studies

The 2012 EU pharmacovigilance legislation made it necessary in some cases to conduct a postauthorization safety study (PASS) in several European countries. The EMA's pharmacovigilance committee, PRAC, and Member State rapporteurs expect some risk management and risk minimization studies to be conducted in such a way. The DSRU has extensive experience designing and implementing event monitoring studies in the UK, but can also work as the coordinating center for PASS studies in a network of several European countries.

The "network of studies" is established by initially identifying research units in multiple European countries that have data sources capable of answering the research question and that can work together. The analysis of the network studies depends on the nature of the study and the expected results. Depending on the characteristics of the data and their analysis,

a metaanalysis or pooling of the results is undertaken. The advantages of these European network studies are many and include conducting PASS studies in several EU countries with their different healthcare systems, increasing the sample size and providing assurance that risk management and risk minimization are working in a number of member states. Such EU network studies provide a way forward to integrating pharmacovigilance across the EU, while complementing the information provided from event monitoring studies within the UK.

Conclusion

Event monitoring contributes to the better understanding of medicines safety. Both signals generated by event monitoring itself and those generated in other systems and studied further by event monitoring have been useful to inform the debates on the safety of medicines, including supporting public health and regulatory decisions.

Like all scientific approaches, event monitoring is evolving, aiming to reduce its weaknesses and enhance its strengths. The most significant development of event monitoring in the last few years has been the introduction of M-PEM and SCEM studies which obtain more information about background history and baseline details of clinical information as well as more details about specific events. M-PEM and SCEM also provide opportunities for comparisons of event rates of different drugs. New methodologic modifications and additions include more effective utilization of information technology, as well as collaborative studies across multiple countries.

Pharmacovigilance and pharmacoepidemiology are emerging and exciting disciplines with evolving study methods. Event monitoring continues to contribute to the progress of these important scientific and public health disciplines.

Acknowledgment from the Director

Event monitoring is a team effort; we are only two members of a large team. The DSRU is most grateful to the thousands of doctors across England who provide the Unit with the safety information which makes its public health work possible. The Unit is equally grateful to the NHSBSA; event monitoring would not be possible without their immense support. We are most grateful to previous and current staff of the DSRU; this chapter is based on their work! Special gratitude goes to Ron Mann for allowing the use of material from the previous editions and to Georgina Spragg who helped in locating research material.

References

1 World Health Organization. *Pharmacovigilance*. www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/ (accessed April 12, 2019).
2 Vargesson N. Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Res* 2015; **105**(2): 140–56.
3 Quibell EP. The thalidomide embryopathy. An analysis from the UK. *Practitioner* 1981; **225**(1355): 721–6.
4 Rawson NSB. *Drug Safety: Problems, Pitfalls and Solutions in Identifying and Evaluating Risk*. Victoria, Canada: Friesen Press, 2016.
5 Medicines Act 1968. www.legislation.gov.uk/ukpga/1968/67 (accessed April 12, 2019).
6 Committee on Safety of Drugs. *Report for 1969 and 1970*. London: HMSO, 1971.
7 Sachs RM, Bortnichak EA. An evaluation of spontaneous adverse drug reaction monitoring systems. *Am J Med* 1986; **81**(5B): 49–55.

8 Abraham J, Davis C. Testing times: the emergence of the practolol disaster and its challenge to British drug regulation in the modern period. *Soc Hist Med* 2006; **19**(1): 127–47.

9 Wright P. Untoward effects associated with practolol administration: oculomucocutaneous syndrome. *BMJ* 1975; **1**(5958): 595–8.

10 Inman WH. Postmarketing surveillance of adverse drug reactions in general practice. I: search for new methods. *BMJ Clin Res Ed* 1981; **282**(6270): 1131–2.

11 Inman WH. Postmarketing surveillance of adverse drug reactions in general practice. II: Prescription-event monitoring at the University of Southampton. *BMJ Clin Res Ed* 1981; **282**(6271): 1216–17.

12 Council for International Organizations of Medical Sciences. *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. Geneva: CIOMS, 2002.

13 Royal College of Physicians of London. *Guidelines on the Practice of Ethics Committees in Medical Research involving Human Subjects*. London: Royal College of Physicians, 1996.

14 General Medical Council. Good Medical Practice. www.gmc-uk.org/static/documents/content/Good_medical_practice_-_English_1215.pdf (accessed April 12, 2019).

15 International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiology Practices. www.pharmacoepi.org/resources/guidelines_08027.cfm (accessed April 12, 2019).

16 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 6). www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE_Rev6.pdf (accessed April 12, 2019).

17 European Medicines Agency. Guidance for the Format and Content of the Protocol of Non-Interventional Post-Authorisation Safety Studies. www.ema.europa.eu/docs/en_GB/document_library/Other/2012/10/WC500133174.pdf (accessed April 12, 2019).

18 International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. *Medical Dictionary for Regulatory Activities (MedDRA)*. www.meddra.org/ (accessed April 12, 2019).

19 Machin D, Campbell M, Fayers P, Pinol A. Table 7.1. In: *Sample Size Tables for Clinical Studies*. Oxford: Blackwell Science, 1997.

20 Machin D, Campbell M, Fayers P, Pinol A. Table 7.2. In: *Sample Size Tables for Clinical Studies*. Oxford: Blackwell Science, 1997.

21 Wilton LV, Stephens MD, Mann RD. Visual field defect associated with vigabatrin: observational cohort study. *BMJ* 1999; **319**(7218): 1165–6.

22 Layton D, Marshall V, Boshier A, Friedmann P, Shakir SAW. Serious skin reactions and selective COX-2 inhibitors: a case series from prescription-event monitoring in England. *Drug Saf* 2006; **29**(8): 687–96.

23 Wilton L, Pearce G, Edet E, Freemantle S, Stephens MD, Mann RD. The safety of finasteride used in benign prostatic hypertrophy: a non-interventional observational cohort study in 14,772 patients. *Br J Urol* 1996; **78**(3): 379–84.

24 Pearce HM, Layton D, Wilton LV, Shakir SAW. Deep vein thrombosis and pulmonary embolism reported in the Prescription Event Monitoring Study of Yasmin. *Br J Clin Pharmacol* 2005; **60**(1): 98–102.

25 Osborne V, Layton D, Perrio M, Wilton L, Shakir SAW. Incidence of venous thromboembolism in users of strontium ranelate: an analysis of data from a prescription-event monitoring study in England. *Drug Saf* 2010; **33**(7): 579–91.

26 Buggy Y, Cornelius V, Fogg C, Kasliwal R, Layton D, Shakir SAW. Neuropsychiatric events with varenicline: a modified prescription-event monitoring study in general practice in England. *Drug Saf* 2013; **36**(7): 521–31.

27 Wilton LV, Pearce GL, Martin RM, Mackay FJ, Mann RD. The outcomes of pregnancy in

women exposed to newly marketed drugs in general practice in England. *Br J Obstet Gynaecol* 1998; **105**(8): 882–9.

28 Osborne V, Davies M, Layton D, Shakir SAW. Utilisation of extended release quetiapine (Seroquel XL™): results from an observational cohort study in England. *Eur Psychiatry* 2016; **33**: 61–7.

29 Beardon PH, McGilchrist MM, McKendrick AD, McDevitt DG, MacDonald TM. Primary non-compliance with prescribed medication in primary care. *BMJ* 199; **307**(6908): 846–8.

30 Finney DJ. The design and logic of a monitor of drug use. *J Chronic Dis* 1965; **18**: 77–98.

31 Heeley E, Riley J, Layton D, Wilton LV, Shakir SA. Prescription-event monitoring and reporting of adverse drug reactions. *Lancet* 2001; **358**(9296): 1872–3.

32 Layton D, Wilton L, Boshier A, Cornelius V, Harris S, Shakir SAW. Comparison of the risk of drowsiness and sedation between levocetirizine and desloratadine: a prescription-event monitoring study in England. *Drug Saf* 2006; **29(**10): 897–909.

33 Acharya NV, Pickering RM, Wilton LW, Shakir SAW. The safety and effectiveness of newer antiepileptics: a comparative postmarketing cohort study. *J Clin Pharmacol* 2005; **45**(4): 385–93.

34 Boshier A, Wilton LV, Shakir SAW. Evaluation of the safety of sildenafil for male erectile dysfunction: experience gained in general practice use in England in 1999. *BJU Int* 2004; **93**(6): 796–801.

35 Flanders WD, Louv WC. The exposure odds ratio in nested case-control studies with competing risks. *Am J Epidemiol* 1986; **124**(4): 684–92.

36 Roy D, Layton D, Marley S, Shakir SAW. Multi-level modelling (MLM) in specialist cohort event monitoring (SCEM) studies. *Pharmacoepidemiol Drug Saf* **26**(S2).

37 Rothman K, Greenland S, Lash T. Validity in epidemiologic studies. In: *Modern Epidemiology*, 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008.

38 McAvoy BR, Kaner EF. General practice postal surveys: a questionnaire too far? *BMJ* 1996; **313**(7059): 732–3; discussion 733–4.

39 Martin RM, Kapoor KV, Wilton LV, Mann RD. Underreporting of suspected adverse drug reactions to newly marketed ("black triangle") drugs in general practice: observational study. *BMJ* 1998; **317**(7151): 119–20.

40 Layton D, Wilton LV, Shakir SAW. Safety profile of celecoxib as used in general practice in England: results of a prescription-event monitoring study. *Eur J Clin Pharmacol* 2004; **60**(7): 489–501.

41 Buggy Y, Cornelius V, Wilton L, Shakir SAW. Risk of depressive episodes with rimonabant. *Drug Saf* 2011; **34**(6): 501–9.

42 Osborne V, Layton D, Fogg C, Tong E, Shakir SAW. Safety of Atrovent(®) CFC-free inhaler: respiratory events reported from an observational cohort study in England. *Int J Pharm Pract* 2018; **26**(4): 356–63.

43 Vlckova V, Cornelius V, Kasliwal R, Wilton L, Shakir S. Hypoglycaemia with pioglitazone: analysis of data from the Prescription-Event Monitoring study. *J Eval Clin Pract* 2010; **16**(6): 1124–8.

44 Layton D, Heeley E, Hughes K, Shakir SAW. Comparison of the incidence rates of thromboembolic events reported for patients prescribed rofecoxib and meloxicam in general practice in England using prescription-event monitoring (PEM) data. *Rheumatology* 2003; **42**(11): 1342–53.

45 Layton D, Harris S, Wilton LV, Shakir SAW. Comparison of incidence rates of cerebrovascular accidents and transient ischaemic attacks in observational cohort studies of patients prescribed risperidone, quetiapine or olanzapine in general practice in England including patients with dementia. *J Psychopharmacol* 2005; **19**(5): 473–82.

46 Boshier A, Wilton LV, Shakir SAW. Evaluation of the safety of bupropion (Zyban) for

smoking cessation from experience gained in general practice use in England in 2000. *Eur J Clin Pharmacol* 2003; **59**(10): 767–73.

47 Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001; **10**(6): 483–6.

48 Mackay FJ, Wilton LV, Pearce GL, Freemantle SN, Mann RD. Safety of long-term lamotrigine in epilepsy. *Epilepsia* 1997; **38**(8): 881–6.

49 Heeley E, Wilton LV, Shakir SAW. Automated signal generation in prescription-event monitoring. *Drug Saf* 2002; **25**(6): 423–32.

50 Layton D, Heeley E, Shakir SAW. Identification and evaluation of a possible signal of exacerbation of colitis during rofecoxib treatment, using Prescription-Event Monitoring data. *J Clin Pharm Ther* 2004; **29**(2): 171–81.

51 Osborne V, Layton D, Shakir SA. Implications of off-label use: an example from the final results of an observational cohort study on Intrinsa® (testosterone patch). *Post Reprod Health* 2014; **20**(2): 48–54.

52 Layton D, Riley J, Wilton LV, Shakir SAW. Safety profile of rofecoxib as used in general practice in England: results of a prescription-event monitoring study. *Br J Clin Pharmacol* 2003; **55**(2): 166–74.

53 Layton D, Souverein PC, Heerdink ER, Shakir SAW, Egberts ACG. Evaluation of risk profiles for gastrointestinal and cardiovascular adverse effects in nonselective NSAID and COX-2 inhibitor users: a cohort study using pharmacy dispensing data in The Netherlands. *Drug Saf* 2008; **31**(2): 143–58.

54 Aurich-Barrera B, Wilton L, Brown D, Shakir S. Paediatric postmarketing pharmacovigilance using prescription-event monitoring: comparison of the adverse event profiles of lamotrigine prescribed to children and adults in England. *Drug Saf* 2010; **33**(9): 751–63.

55 Osborne V, Davies M, Layton D, Shakir SAW. Utilisation and safety of deferasirox: results from an observational cohort study in England. *Drug Saf* 2018; **41**(3): 267–75.

56 Martin RM, Biswas PN, Freemantle SN, Pearce GL, Mann RD. Age and sex distribution of suspected adverse drug reactions to newly marketed drugs in general practice in England: analysis of 48 cohort studies. *Br J Clin Pharmacol* 1998; **46**(5): 505–11.

57 Aurich-Barrera B, Wilton LV, Shakir SAW. Use and risk management of carvedilol for the treatment of heart failure in the community in England: results from a modified prescription-event monitoring study. *Drug Saf* 2009; **32**(1): 43–54.

58 McNaughton R, Lynn E, Osborne V, Coughtrie A, Layton D, Shakir S. Safety of Intranasal quadrivalent live attenuated influenza vaccine (QLAIV) in children and adolescents: a pilot prospective cohort study in England. *Drug Saf* 2016; **39**(4): 323–33.

59 Mackenzie IS, MacDonald TM, Shakir S, et al. Influenza H1N1 (swine flu) vaccination: a safety surveillance feasibility study using self-reporting of serious adverse events and pregnancy outcomes. *Br J Clin Pharmacol* 2012; **73**(5): 801–11.

60 Perrio MJ, Wilton LV, Shakir SAW. A modified prescription-event monitoring study to assess the introduction of Flixotide Evohaler into general practice in England: an example of pharmacovigilance planning and risk monitoring. *Pharmacoepidemiol Drug Saf* 2007; **16**(9): 969–78.

61 Perrio MJ, Wilton LV, Shakir SAW. A modified prescription-event monitoring study to assess the introduction of Seretide Evohaler in England: an example of studying risk monitoring in pharmacovigilance. *Drug Saf* 2007; **30**(8): 681–95.

62 Davies M, Coughtrie A, Layton D, Shakir SAS. Use of atomoxetine and suicidal ideation in children and adolescents: results of an observational cohort study within general practice in England. *Eur Psychiatry* 2017; **39**(Suppl C): 11–16.

16

Primary Data Collection for Pharmacoepidemiology

Nancy A. Dreyer[1], *Ana Filipa Macedo*[2], *and Priscilla Velentgas*[1]

[1] *IQVIA, Real-World Insights, Cambridge, MA, USA*
[2] *QuintilesIMS, Madrid, Spain*

Introduction

Primary data collection refers to data that are collected specifically for a given research study or program, while secondary data are data that were collected to meet needs other than the research for which they are being used. Primary data collection can be used in all types of pharmacoepidemiologic study designs, both interventional and noninterventional, cohort, and case–control studies, as well as patient registries, which are conceptually a data collection structure for disease- or product-related studies, and in practice may closely resemble observational cohort study designs. These studies generally address effectiveness and safety of various medical treatments, and also are used to characterize diseases including progression over time.

Research Questions That Require Primary Data

The nature of the research question(s) and accompanying study design determine the need for collection of primary data from clinicians, patients, and/or others to address the study's aims, and are generally weighed against the sufficiency of availability of the required information in existing data sources (see Chapter 17). Some studies may combine existing data with limited supplementary primary data collection in order to collect critical aspects of the patient or healthcare provider experience; in those situations, accurate linkage is required between primary data collection and existing data (see section on "Hybrid or Enriched Designs" later in this chapter).

Research questions that may require primary data include the following.

Designs Involving Randomization

Pragmatic and explanatory randomized clinical trial designs (see Chapter 32) generally necessitate at least minimal site and/or patient contact for screening to determine eligibility and consent to participate in the trial. Pragmatic randomized trials are particularly useful for studying products that may not be widely used, either because they are newly approved or because they are not covered by health insurance plans or their costs are only covered to a small extent. Assignment of treatment through a randomization schedule assures that the product will be

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

used in the target population of interest, and in the doses, sequence, and/or combinations of interest. Treatment may or may not be blinded, depending on whether the outcome can be measured objectively. Blinding (or masking) treatment is particularly important for outcomes that may be heavily influenced by the patient's or clinician's estimation of the benefit of the study treatment [1]. Placebos are used in drug development but in pragmatic randomized trials (see Chapter 26), the comparators are more frequently either a single product or whatever is being used in the locale of interest as treatment for the condition of interest, often called the "standard of care."

While randomized studies frequently involve extensive primary data collection to meet the trial objectives, these considerations overlap with those of observational designs and will be addressed further in that context.

Endpoint Assessment/Adjudication

Studies may require collection of detailed primary data for endpoint assessment and/or for endpoint adjudication. Collection of primary data for endpoint assessment becomes important when study endpoints are not consistently recorded in available medical records systems, or are not recorded with the reliability, timing or frequency needed to meet study aims. Additional primary data may also be collected to validate or confirm endpoints collected through secondary data or patient self-reports.

Endpoint validation can be particularly important for studies that use the patient as the primary reporter, since patients report as consumers, and clinical validation may be needed to confirm the endpoint of interest. For example, in the European PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics) Consortium, funded by the Innovative Medicines Initiative, data were collected directly from pregnant women recruited online from the UK, Denmark, The Netherlands, and Poland [2]. Researchers learned that women could accurately report serious birth defects, but there were many reports of potential abnormalities that were difficult to classify without more clinical information [3].

Some observational as well as interventional studies, especially those designed to meet postmarketing requirements with safety or effectiveness endpoints that require an additional level of rigor, may include a full or modified approach to clinical endpoint review or adjudication by a central committee over and above the reporting by individual study sites. Reasons for this additional adjudication include concerns regarding investigator bias, if they hold strong opinions as to the benefit or harm associated with a treatment under study, the need to apply consistent standard definitions given variability in diagnostic criteria in usual practice, and lack of detail or inconsistent use of standard coding practices in accessible secondary data sources [4].

Clinical Assessments Not Consistently Captured in Secondary Data

Even as the collection of medical data from routine care is increasingly recorded and available from electronic as well as paper medical records, substantial variability in performing assessments on the part of healthcare providers according to individual and local practice, magnified by patients' variability in coming in for recommended routine visits, limits the extent to which clinical data from secondary sources can be used to address some research questions. See section on "Clinician- or Site-Reported Outcomes" later in this chapter.

Commonly, prospective studies, including planned analysis of laboratory data or imaging studies over time, may incorporate primary data collection to ensure complete collection of assessments and that timing of assessments is aligned with the study follow-up period. Additionally, use of a central lab to reduce variability in laboratory measures may be

considered to further increase the validity of study results. The PROVALID study (PROspective cohort study in patients with type 2 diabetes mellitus for VALIDation of biomarkers), launched in five EU countries (Austria, Hungary, The Netherlands, Poland, and Scotland), will obtain laboratory measurements on 4000 enrolled patients with type 2 diabetes treated in the primary care setting to examine the impact of medication and predict the clinical course of disease, including renal and cardiovascular events. PROVALID requires collection of a minimum set of clinical data parameters, with the option to collect many additional laboratory and medical history characteristics on enrolled patients [5].

Characterization of Patient-Reported Outcomes Not Captured in Secondary Data

It is often important to evaluate the burden of disease on patients and how that burden is affected by various treatments. For example, the symptom burden on work-related abilities was evaluated in patients being treated for locally recurrent or metastatic breast cancer. Information on the ability to perform work was ascertained directly from patients [6]. More information on patient engagement and direct-to-patient designs can be found in an electronic book on 21st Century Patient Registries [7] and in the section on "Patient-Reported Outcomes and Other Patient-Reported Measures" later in this chapter.

Studies of Rare Populations

When it is necessary to assemble as large and representative a sample as possible from a rare population, one or more existing data sources may not capture enough of the patient population of interest to address study aims.

Rare disease registries and pregnancy exposure registries commonly face the challenge of a small number of patients with the condition or exposure of interest, distributed over many countries, for which no single existing data source likely includes enough patients to address research aims.

The lysosomal storage disorders (LSDs) are a group of genetic conditions characterized by enzyme deficiencies that leave cells unable to clear waste products, which accumulate with a range of harmful physical effects if not successfully treated. A number of global registries developed to study the natural history, treatment patterns, and effectiveness of existing treatments for LSDs, including Fabry, Gaucher, hereditary angioedema, mucopolysaccharidosis I (MPS 1) and 2 (MPS 2, Hunter syndrome), and Pompe disease have been sponsored by manufacturers of enzyme replacement therapies and other treatments, and by patient associations [8–12]. In the case of such rare conditions, where small numbers of patients and the providers who treat them are scattered across the globe, a registry serves multiple purposes of linking individual patients with each other and with a treatment community, as well as potentially drawing upon a large and representative population available for research. The validity of conclusions that may be drawn from analyses of this kind of registry data depends on the degree to which subjects in the registry are not selectively included because of their treatment outcomes, and that follow-up is relatively complete.

Vaccine Safety

Some adverse events (AEs) of interest may not be routinely captured in existing data sources (see Chapter 20).

Beginning in 2014, the European Medicines Agency (EMA) has required annual enhanced safety surveillance (ESS) for all seasonal influenza vaccines. The interim guidance from the Pharmacovigilance Risk Assessment Committee (PRAC) included the requirement to collect data that would support the identification of any significant change in frequency or severity of reactogenicity in comparison with previous

years' experience with the same vaccine composition [13]. Reactogenicity AEs of interest as specified in the guidance include vaccination site reactions, headache, malaise, myalgia, shivering, rash, vomiting, nausea, arthralgia, decreased appetite, irritability, and crying (in pediatric vaccinees less than 5 years of age). Such symptoms, especially when mild or moderate in severity, are not commonly reported to healthcare providers and thus medical records are not a useful source of data to study these outcomes following vaccination. Instead, studies and surveillance activities to implement the requirements for ESS have incorporated patient (and proxy for pediatric patients) self-report of occurrence of symptoms to obtain this information in a systematic manner, for example through distribution of Safety Report Cards allowing patients to report these symptoms by telephone or mail [14].

Studies of Medical Devices

Device studies may require information on batch and manufacture location that are not often available in existing data sources (see Chapter 21). Additional information on the "operator" or healthcare provider implanting or administering the device may also help to provide a full characterization of product safety and effectiveness.

For example, the National Cardiovascular Data Registry compares the effectiveness of drug-eluting and bare metal stents in reducing risk of death or myocardial infarction [15]. Studies of the Björk–Shiley heart valve showed the importance of manufacturing site as a risk factor for valve failure [16].

Special Requirements, Controlled Distribution Products

Some products are approved by regulatory authorities with special requirements for reporting mainly related to safety concerns. Often such requirements necessitate data collection specific to the product to ensure robust monitoring of safety concerns.

Several active mandated safety registries and an example of a multisponsor pediatric safety registry were described in the Pink Sheet in 2009, "Registries Rising: FDA Looking at TNF Inhibitors; AHRQ Updates Standards" [17]. These examples included a pregnancy registry for the same product (ClinicalTrials.gov identifier NCT01026077), a pregnancy registry as part of a restricted distribution program for eltrombopag (ClinicalTrials.gov identifier NCT01064336), a thrombopoietin receptor agonist for treatment of idiopathic thrombocytopenia purpura (ITP) sponsored by GlaxoSmithKline, and a safety registry for teriparatide, an anabolic treatment for osteoporosis and an expanded indication of glucocorticoid-induced osteoporosis sponsored by Eli Lilly.

Hybrid or Enriched Designs

Some studies may not rely on primary or secondary data alone. The terms "hybrid" or "enriched" are frequently used to describe study designs that draw upon both primary and secondary data, with some data collected *de novo* specifically for the purposes of the study and other study-specific data collected via probabilistic or deterministic linkage with other data sources, such as electronic health records, administrative claims and billing data, vital records, and genetic information.

The DISCOVER study is an example of an enriched study. The study objective is to characterize and describe the management of patients initiating second-line therapy for type 2 diabetes; data are being collected from patients at sites in 38 countries, with linkage of electronic health records where feasible. Information is collected about diabetes management including patients' health-related quality of life [18]. For more information see www.discoverdiabetes.com/.

Description

Clinician- or Site-Reported Outcomes

The traditional approach to primary data collection has been to recruit healthcare providers (HCPs) for study and then for those HCPs to recruit patients following agreed-upon inclusion and exclusion criteria as described in the protocol. When HCPs can form accurate assessments after observation of a patient's health condition, those clinician-reported outcomes (ClinROs) can become the measurement on which a study endpoint is based. Traditionally, data collection was performed on paper, but now most such data collection is electronic through case report forms (CRFs) designed specifically for study purposes. This model requires institutional board review (ethical review) and site contracts with each investigator. Generally, investigators expect some payment for data collection, and such payments must be proportional to time spent and fair market value.

Patient-Reported Outcomes and Other Patient-Reported Measures

There is a growing interest in collecting data from patients on exposures and outcomes. These data can be collected directly from patients without the intervention of HCPs or can be collected from patients during a study visit or electronically in between study visits.

Recent research has established the validity of patient-reported prescription medication use, laying a foundation for its reliability as well as demonstrating the rich additional information that can be obtained, such as nonprescription medication use, recreational drug use, smoking, and alcohol intake [3]. There are a large number of validated patient-reported outcome (PRO) measures that can provide important insights into the patient experience, including treatment satisfaction, quality of life, ability to care for oneself, work, etc., and new tools are constantly in development. If patients will be contacted for any study data, supplementary information can often be obtained at a small marginal cost, especially if data collection is as brief as possible.

Despite good motivations, the more data that are sought from patients, the less likely it is that they will complete the questions or continue to participate in the study. Also, clinical input may be needed to accurately and fully report patients' medical histories and serious health events of special interest. Clinical validation is often obtained for patient-reported clinical events of special interest.

Objective measures of patients' physical activity are of interest in numerous therapeutic areas for pharmacoepidemiologic research, including chronic obstructive pulmonary disease, cardiovascular disease, and depression. Quantification of the validity and reliability of wearable sensors that collect information about physical activity and various other clinically useful data will encourage greater use in longitudinal observational studies and pragmatic trials. At the time of this chapter, mainly smaller or cross-sectional observational studies and a few randomized clinical trials have incorporated accelerometry measurement of physical activity as a study exposure, covariate of interest, or endpoint [19,20].

Registries as Means of Data Collection

As previously mentioned, registries may be established to fill a need for data collection that may support multiple research and/or public health surveillance objectives.

Population-based state, regional, and national cancer registries have played a major role in cancer surveillance, by quantifying cancer incidence and mortality, and trends over time throughout the world, and in pharmacoepidemiology, by providing data on prognostic factors, treatment, and outcomes for analysis within single or across linked databases. In the United States, the Surveillance, Epidemiology,

and End Results (SEER) program of the National Cancer Institute works from a network of 18 cancer registries in 14 states that actively collect information on all reported cancers diagnosed in their coverage areas [21].

Pharmacoepidemiologic studies using cancer registries have included case–control studies such as the Cancer and Steroid Hormone (CASH) study of oral contraceptive use and breast [22], ovarian [23], and endometrial [24] cancer, and patterns of care studies of the dissemination of advanced cancer treatment modalities throughout different population groups and into community practice [25–27]. With approval, researchers may be granted access to the SEER-Medicare linked data files, which include Medicare claims before, during, and following cancer diagnosis and treatment [28]. Topics studied include influences of treatment, facility, and provider characteristics and interventions on survival and cost outcomes [29–31], as well as disparities in care [32].

Biobanks/Specimen Banks

Clinical data are increasingly being linked with biorepository data to guide researchers in the identification of biomarkers that are predictive of clinical outcomes and support the development of targeted therapies, for example by identifying patients whose tumors harbor a genomic variant that can potentially be targeted by a new drug. Some biobanks are being set up internationally, such as the EuroBioBank network, which was created to support research on rare diseases [33]. The UK Biobank is also a good example of an international long-term registry accessible for research, which is following around 500 000 volunteers for at least 25 years to investigate the contributions of genetic predisposition and environmental exposure (including nutrition and lifestyle) to disease development, and gain valuable insights to support the development of new medicines [34].

Guidelines on the Quality of Data Collection

There are a variety of recent guidelines that broadly speak to studies that use primary data collection, such as the Guidelines for Good Pharmacoepidemiologic Practice developed by the International Society for Pharmacoepidemiology [35], the checklist for study protocols developed by the European Network of Centers for Pharmacoepidemiology and Pharmacovigilance [36], and principles of good epidemiologic methods and practice (see Chapter 3) [37]. Also, the GRACE (Good Research for Comparative Effectiveness) Principles for conducting and evaluating observational studies of comparative effectiveness are applicable to studies that use primary data collection and may help guide the study design [38,39]. More detail about operational aspects of primary data collection can be found in the Registries for Evaluating Patient Outcomes: A User's Guide [40]. In that guide, the chapter on assessing quality provides a detailed listing of various aspects of primary data collection relating to research quality and evidence quality, describing both the basic elements of good practice and potential enhancements [41].

Strengths

A notable strength of primary data collection is that it can address research objectives that require information that is not otherwise accessible or not consistently recorded in available secondary data sources. This type of information can be particularly meaningful to clinicians, patients, regulators, payers, and those involved in drug development, and is often more directly useful than inferences derived from billing data and often cryptic and spotty electronic health data.

Studies that use data collection directly from patients also provide the opportunity to follow

patients over long periods of time and to evaluate a variety of outcomes. For chronic diseases, patients may be followed for years by their treating physicians, regardless of whether a patient's health insurance program changes – an important limitation of health insurance claims data. Patients may also be followed directly for self-reported outcomes, both for clinical outcomes (which may be confirmed with medical records or by physicians, as needed) and quality-of-life measures. It is often the patient's relationship with the physician, and the physician's relationship with the research program, that is particularly effective for long-term retention. For example, the VIRGO study followed patients with metastatic breast cancer in collaboration with their treating physicians to understand the impact of treatment on quality of life as the disease progressed [42].

Limitations

Primary data collection requires cooperation of data contributors, often over long periods of time (follow-up). While it would seem that altruism should be a sufficient motivation, experience has shown that successful primary data collection requires an infrastructure supporting patient and/or physician enrollment and retention, as well as an active program of data curation to assure that the data collected are accurate and reliable. Further, one must be mindful about the use of patient-centered endpoints, especially pertaining to general and disease-specific quality of life assessments and to detailed information on past exposures, such as in case–control studies where the study outcome is known to the patients at the time of the assessment [43]. Like all data, the contribution to be made by patients in recall of past medical diagnoses and exposures of interest to pharmacoepidemiologic studies must be considered carefully in view of their strengths and limitations [44–46].

It is also important to keep in mind that consumers are likely to report serious AEs quite differently from clinicians, and such reports may require clinical validation for use for research purposes. For example, the PROTECT study examined medication use during pregnancy using biweekly or monthly questionnaires administered via the internet, with the frequency of follow-up determined according to the participants' choosing [47]. Respondents were able to clearly describe serious but rare birth defects, but used common terms that were difficult to code for less serious conditions that might not meet the clinician threshold for being considered a birth defect, for example the eyes were too small or the nose appeared to be off-center.

In addition to data quality, the two major challenges in primary data collection relate to enrollment and retention. It is often difficult to find and recruit clinicians who treat patients of interest, or patients who have the exposures or conditions of interest, no matter how important the research question. For example, Pfizer made great fanfare over the launch of its first "virtual trial" called REMOTE, that focused on overactive bladder treatment. The study was halted less than a year after launch because of poor recruitment, with a spokesman from the sponsoring organization commenting that social media was a great way to spread awareness of the trial but did not build enough trust for patients to actually sign up [48].

"Sufficient" retention (regardless of the target) can be difficult to achieve when using primary data collection. Retention rates are often higher for studies that are (1) responsive to the needs of patients and physicians. so they are motivated to continue participating (a special concern for pregnancy registries and other vulnerable populations), and (2) parsimonious in their data collection [49]. Operational challenges relate to the need to deploy primary data collection systems that are easy to use, and simple enough to encourage steady reporting but

which do not result in reporting fatigue. Multiple methods of data entry such as internet, text messaging, and/or mail can be an advantage, considering the demographics of the target population, although many researchers believe that there is a positive impact on retention from consistent personal interactions between study staff and clinical investigators and/or patients, a technique often deployed in pregnancy registries.

While there is some optimism that newer technologies and the nearly universal adoption of smartphones would support the use of text messaging and internet-based patient surveys, results from the PROTECT study raised a cautionary note [47]. Researchers noted that internet-based recruitment of pregnant women was surprisingly difficult and study retention was low, speculating that although it was relatively easy to send questionnaires frequently, participants appeared to tire quickly of responding to the same questions over time.

Enrollment logs that record some information about those who are eligible to participate but decline or later drop out can be used to evaluate selection bias and the impact of loss to follow-up. For example, studies that collect some basic personal identifiers such as name (first, last, and middle initial) and date of birth can be linked with the National Death Index in the United States to search for deaths and obtain information on cause of death.

Particular Applications

To provide further illustration of some of the applications of primary data collection in modern pharmacoepidemiologic research, several examples are described in further detail in this section. These include a prospective comparative effectiveness research (CER) study incorporating collection of clinical endpoints and PROs, a novel hybrid study intended to provide data in support of a label expansion with FDA, use of large registry data as a framework for conduct of multiple observational studies, and incorporation of measures of physical activity through accelerometry as part of the UK Biobank effort.

The Registry in Glaucoma Outcomes Research (RiGOR) study, funded by the US Agency for Healthcare Research and Quality, was a prospective observational study that used primary data collection to address which treatment strategy for open-angle glaucoma was associated with the greatest improvement in patient outcomes [50]. The study found that patients treated with incisional surgery after failing at least one course of medication were twice as likely as patients treated with additional medication to achieve a 15% reduction in intraocular pressure (IOP) at 12 months, while patients treated with laser surgery had similar results to those who were treated with additional medication [51]. While IOP is routinely recorded when glaucoma patients see their ophthalmologists, in order to ensure complete assessment of IOP at around six and 12 months of follow-up, it was a required element in the study's CRF, along with a vast array of other detailed clinical information. The RiGOR study also included several validated PROs assessments as secondary endpoints, which further required patients to complete these questionnaires at the time of a study visit or at home through mail or electronic means [52].

The Bioventus Exogen® device registry is a novel example of a hybrid or enriched design involving both primary and secondary data. This study was planned following extensive discussions with the US Food and Drug Administration Center for Drug Evaluation and Research (FDA CDER) regarding this novel design for a label expansion. The study utilizes a prospective direct-to-patient product registry linked with a propensity score matched comparator group from a commercial claims database for study of a device used to treat bone fracture nonunion, currently used broadly outside of labeled indications to treat fracture [53].

Two European cancer registries provide examples of registry infrastructure being built to address innumerable current and future research questions pertaining to cancer incidence and survival. The EUROCARE (European Cancer Registry Based Study on Survival and Care of Cancer Patients) registry is a very large collaborative research project on cancer survival [54]. The registry started in 1989, aiming to provide updated descriptions of cancer survival time trends and differences across European countries, to measure cancer prevalence, and study patterns of care of cancer patients. Its fifth and current edition, EUROCARE-5, includes data on more than 21 million cancer diagnoses provided by 116 cancer registries in 30 European countries [55]. At least 165 publications have been generated from EUROCARE-1–5, covering trends in survival across a very broad range of cancer types as well as patterns of care and predictors of survival and other outcomes [56]. The European Cancer Observatory (ECO) is another important project developed at the International Agency for Research on Cancer (IARC) in partnership with the European Network of Cancer Registries (ENCR) in the framework of the EUROCOURSE project which is supported by the European Commission. It presents national estimates of cancer incidence, mortality, and prevalence for 24 major cancer types in 40 European countries for 2012. This registry also allows for the exploration of geographical patterns and temporal trends [57]. Access to cancer datasets is possible on a user registration and permission request basis [58].

An additional area of data collection for the UK Biobank to support the inclusion of objective measures of physical activity in large-scale observational studies has been to obtain measures of physical activity from accelerometers from over 100 000 participants [59]. Forty-five percent of those invited to wear accelerometers for seven days accepted the invitations; from these, over 93% provided sufficient valid data for analysis.

The Future

Newer methods of collecting data directly from patients and from doctors, including internet-based data collection, integrated voice response systems with auto-coding features, and use of wearable sensors, will facilitate rapid accumulation of data; however, the validity, accuracy, and usefulness of these new methods need to be evaluated. Factors that affect recruitment and retention also must be considered. Just because it may be relatively easy to contribute data does not mean that patients will continue to do so over a long period of time. In focus groups used in the PROTECT study, pregnant women reported that some reimbursement for their time for each survey completed, even something nominal, would increase their interest in staying in a study since it would show that researchers value their input [60].

An exciting new direction is the potential for enriched studies to be conducted on a more routine basis. For example, there is growing interest in expanding patient registry data collection to allow patients to authorize linkage with other of their healthcare data on an ongoing basis. In these situations and where local law allows, the Informed Consent document is likely to require inclusion of general information described below.

- I have read the contents of this form and I understand, agree, and allow my Health Plan, The REGISTRY Coordinating Center at XXXX, and my healthcare provider to use and release information about me as described above. I also understand that signing this form is of my own free will and will in no way affect the medical care that I receive.
- If I choose not to participate in this linkage to claims information, I understand that my Health Plan will not base decisions regarding my treatment, eligibility for benefits, enrollment in a Health Plan, or payment of claims on my decision regarding study participation.

- If I choose not to participate in this linkage it will not impact my participation in the main REGISTRY. Additionally, I can choose to withdraw from the linkage at any time and remain in the main REGISTRY if I choose.
- I have the right to withdraw this approval at any time by giving written notice of my withdrawal to the REGISTRY. I understand that my withdrawing this approval will not affect any action taken before I do so. I also understand that once my information is shared outside the REGISTRY, it will not be traceable to me. However, the information from my records will still be protected by other privacy rules and agreements.
- At my request, I will be given a copy of this form either when I sign it or while the study is ongoing.

In summary, despite the growing availability of large amounts of data on treatments and patients' clinical experience, it is unlikely that such data will ever contain all information needed for every study purpose; thus, the need for primary data collection will remain. Traditionally, HCPs have been the primary reporters/recorders of data for studies that use primary data collection, although there is growing interest in collecting data directly from patients either exclusively or in combination with wearable devices and/or clinician-reported data, and/or to enrich existing data. While the methods for such data collection can and will change over time, it is likely that researchers will always need to invest time in data curation to assure that the data are accurately represented and to check for critical data that are systematically missing. Primary data collection will continue to be a mainstay of pharmacoepidemiologic research, either as the sole method of data collection or as a key component of research that uses many modes of data collection.

References

1 Dreyer NA. Advancing a framework for regulatory use of real-world evidence: when real is reliable. *Ther Innov Regul Sci* 2018; **52**(3): 362–8.

2 www.imi-protect.eu (accessed April 17, 2019).

3 Dreyer NA, Blackburn SCF, Mt-Isa S, *et al*. Direct-to-patient research: piloting a new approach to understanding drug safety during pregnancy. *JMIR Public Health Surveill* 2015; **1**(2): e22.

4 Seltzer JH, Heise T, Carson P, *et al*. Use of endpoint adjudication to improve the quality and validity of endpoint assessment for medical device development and post marketing evaluation: rationale and best practices. A report from the Cardiac Safety Research Consortium. *Am Heart J* 2017; **190**: 76–85.

5 Eder S, Leirer J, Kershbaum J, *et al*. A Prospective Cohort Study in Patients with Type 2 Diabetes Mellitus for Validation of Biomarkers (PROVALID) – study design and baseline characteristics. *Kidney Blood Press Res* 2018; **43**: 181–90.

6 Cleeland CS, Mayer M, Dreyer NA, *et al*. Impact of symptom burden on work-related abilities in patients with locally recurrent or metastatic breast cancer: results from a substudy of the VIRGO observational cohort study. *Breast* 2014; **23**(6): 763–9.

7 Gliklich RE, Dreyer NA, Levy M, Christian JB (eds). *21st Century Patient Registries. Registries for Evaluating Patient Outcomes: A User's Guide. Third Edition Addendum (Overview)*. Rockville, MD: Agency for Healthcare Research and Quality, 2018. www.effectivehealthcare.ahrq.gov/topics/registries-guide-3rd-edition-addendum/overview (accessed April 17, 2019).

8 Weinreb N, Kaplan P. The history and accomplishments of the ICGG Gaucher Registry. *Am J Hematol* 2015; **90**(Suppl 1): S2–5.

9 Germain DP, Charrow J, Desnick RJ, *et al*. Ten-year outcome of enzyme replacement therapy with agalsidase beta in patients with Fabry Disease. *J Med Genet* 2015; **52**(5): 353–8.

10 United States Hereditary Angioedema Association. US HAE Association Scientific Registry: Patient Driven Research for a Cure. www.haea.org/resources/us-haea-scientific-registry/ (accessed April 17, 2019)

11 D'Aco K, Underhill L, Rangashari L, *et al*. Diagnosis and treatment trends in mucopolysaccharidosis I: findings from the MPS I Registry. *Eur J Pediatr* 2012; **171**(6): 911–19.

12 Güngör D, Kruijshaar ME, Plug I, *et al*. Impact of enzyme replacement therapy on survival in adults with Pompe disease: results from a prospective international observational study. *Orphanet J Rare Dis* 2013; **8**: 49.

13 Pharmacovigilance Risk Assessment Committee (PRAC). *Interim Guidance on Enhanced Safety Surveillance for Seasonal Influenza Vaccines in the EU*. London: European Medicines Agency, 2014. www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/04/WC500165492.pdf (accessed April 17, 2019).

14 Bricout H, Chabanon AL, Souverain A, Sadorge C, Vesikari T, Caroe T. Passive enhanced safety surveillance for Vaxigrip and Intanza 15 µg in the United Kingdom and Finland during the northern hemisphere influenza season 2015/16. *Euro Surveill* 2017; **22**(18): 30527.

15 Douglas PS, Brennan JM, Anstrom KJ, *et al*. Clinical effectiveness of coronary stents in the elderly: results from 262,700 Medicare patients in ACC-NCDR. *J Am Coll Cardiol* 2009; **53**(18): 1629–41.

16 Walker AM, Funch DP, Sulsky SI, Dreyer NA. Manufacturing characteristics associated with strut fracture in Björk–Shiley 60° convexo-concave heart valves. *J Heart Valve Dis* 1995; **6**(4): 640–8.

17 Hobbs MN. Registries Rising: FDA Looking At TNF Inhibitors; AHRQ Updates Standards. *Pink Sheet* 2009; **71**(34): 8–9.

18 www.discoverdiabetes.com/DISCOVER-study.html (accessed April 17, 2019).

19 Minakata Y, Morishida Y, Ichikawa T, *et al*. Effects of pharmacologic treatment based on airflow limitation and breathlessness on daily physical activity in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2015; **10**: 1275–82.

20 Watz H, Troosters T, Beeh KM, *et al*. ACTIVATE: the effect of aclidinium/formoterol on hyperinflation, exercise capacity, and physical activity in patients with COPD. *Int J Chron Obstruct Pulmon Dis* 2017; **12**: 2545–58.

21 National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Overview of the SEER Program. http://seer.cancer.gov/about/ (accessed April 17, 2019)

22 Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. Oral contraceptive use and risk of breast cancer. *N Engl J Med* 1986; **315**: 405–11.

23 No authors listed. Oral contraceptive use and the risk of ovarian cancer. *JAMA* 1983; **249**: 1596–9.

24 No authors listed. Oral contraceptive use and the risk of endometrial cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. *JAMA* 1983; **249**: 1600–4.

25 Shavers VL, Harlan LC, Jackson M, Robinson J. Racial/ethnic patterns of care for pancreatic cancer. *J Palliat Med* 2009; **12**(7): 623–30.

26 Stroup AM, Harlan LC, Trimble EL. Demographic, clinical, and treatment trends among women diagnosed with vulvar cancer in the United States. *Gynecol Oncol* 2008; **108**(3): 577–83.

27 Mariotto AB, Feuer EJ, Harlan LC, Abrams J. Dissemination of adjuvant multiagent

chemotherapy and tamoxifen for breast cancer in the United States using estrogen receptor information: 1975–1999. *J Natl Cancer Inst Monogr* 2006; **36**: 7–15.

28 National Cancer Institute. Surveillance, Epidemiology, and End Results Program. SEER-Medicare Linked Database. https://seer.cancer.gov/ (accessed April 17, 2019)

29 Cheung WY, Neville BA, Earle CC. Etiology of delays in the initiation of adjuvant chemotherapy and their impact on outcomes for Stage II and III rectal cancer. *Dis Colon Rectum* 2009; **52**(6): 1054–63; discussion 1064.

30 Hershman DL, Buono D, McBride RB, Tsai WY, Neugut AI. Influence of private practice setting and physician characteristics on the use of breast cancer adjuvant chemotherapy for elderly women. *Cancer* 2009; **115**(17): 3848–57.

31 Breslin TM, Morris AM, Gu N, *et al*. Hospital factors and racial disparities in mortality after surgery for breast and colon cancer. *J Clin Oncol* 2009; **27**(24): 3945–50.

32 Rouhani P, Arheart KL, Kirsner RS. Differences in melanoma outcomes among Hispanic Medicare enrollees. *J Am Acad Dermatol* 2010; **62**(5): 768–76.

33 EuroBioBank Network. www.eurobiobank.org/ (accessed April 17, 2019).

34 UK Biobank. www.ukbiobank.ac.uk/ (accessed Aapril 17, 2019).

35 Epstein M. Guidelines for Good Pharmacoepidemiology Practices (GPP). *Pharmacoepidemiol Drug Saf* 2015; **25**: 1099–557.

36 www.encepp.eu/standards_and_guidances/checkListProtocols.shtml (accessed April 17, 2019).

37 Rothman KH, Greenland S, Lash TL. *Modern Epidemiology*, 3rd edn. New York: Lippincott Williams and Wilkins, 2008.

38 Dreyer NA, Schneeweiss S, McNeil B, et al., on behalf of the GRACE Initiative. GRACE Principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010; **16**(6): 467–71.

39 Dreyer NA, Bryant A, Velentgas P. The GRACE Checklist: a validated assessment tool for high quality observational studies of comparative effectiveness. *J Manage Care Specialty Pharm* 2016; **22**(10): 1107–13.

40 Gliklich R, Dreyer N, Leavy M (eds). *Registries for Evaluating Patient Outcomes: A User's Guide*, 4th edn. Rockville, MD: Agency for Healthcare Research and Quality, 2019.

41 Dreyer NA, Gliklich RE, Leavy MB. Assessing quality. In: *Registries for Evaluating Patient Outcomes: A User's Guide*, 4th edn (eds R. Gliklich, N. Dreyer, M. Leavy). Rockville, MD: Agency for Healthcare Research and Quality, 2019, pp. 373–89.

42 Kaufman PA, Dreyer NA, Mayer M, *et al*. Capturing breast cancer patients experience beyond disease progression: implementation of a patient-reported outcome (PRO) substudy in the VIRGO metastatic breast cancer observational cohort study. *J Clin Oncol* 2010; **28**: 6138.

43 Johansen C. Study designs may influence results: the problem with case-control studies on the epidemiology of glioma. *Br J Cancer* 2017; **116**: 841–8.

44 Laursen N, Bourke A, Dreyer NA, Blackburn SF. Can patients accurately report their drugs? A comparison of self-reported medication with a national register in Demark in a subset of the PROTECT Pregnancy Cohort. Poster presented at the International Society of Pharmacoepidemiology & Drug Safety Annual Conference, 2018.

45 Laws MB, Lee Y, Taubin T, *et al*. Factors associated with patient recall of key information in ambulatory specialty care visits: results of an innovative methodology. *PLoS One* 2018; **13**(2): e0191940.

46 Sunderman AC, Hartmann KE, Jones SH, *et al*. Validation of maternal recall of early pregnancy medication exposure using

prospective diary data. *Ann Epidemiol* 2017; **27**(2): 135–9.e2.

47 Dreyer NA, Blackburn SCF, Hliva V, *et al*. Balancing the interests of patient data protection and medication safety monitoring in a public-private partnership. *JMIR Med Inform* 2015; **3**(2): e18.

48 www.clinicalleader.com/doc/pfizer-perseveres-in-pioneering-virtual-clinical-trials-0001 (accessed April 17, 2019).

49 Mendelsohn AB, Dreyer NA, Mattox PW, *et al*. Characterization of missing data in clinical registry studies. *Ther Innov Regul Sci* 2014; **48**(5): 574–82.

50 Coleman A, Lum F, Velentgas P, *et al*. RiGOR: a prospective observational study comparing the effectiveness of treatment strategies for open-angle glaucoma. *J Comp Effectiveness Res* 2016: **5**(1): 65–78.

51 Coleman A, Lum F, Velentgas P, *et al*. Impact of treatment strategies for open angle glaucoma on intraocular pressure: the RiGOR study. *J Comp Effectiveness Res* 2016; **5**(1): 87–98.

52 Coleman AC, Lum F, Gliklich R, Velentgas P, Su Z. Quality of life and visual acuity outcomes in the Registry in Glaucoma Outcomes Research Study 2016. https://doi.org/10.2217/cer.15.59 (accessed April 17, 2019).

53 Mack C, Pavesio A, Kelly K, *et al*. Breaking news: study of fracture healing in patients at risk of non-union. Poster presented at the 33rd International Conference on Pharmacoepidemiology & Therapeutic Risk Management, Montreal, Canada (23–27 August, 2017).

54 World Health Organization, International Agency for Research on Cancer. Eurocare study: survival of cancer patients in Europe. www.eurocare.it/Home/tabid/36/Default.aspx (accessed April 17, 2019).

55 Minicozzi P, Otter R, Primic-Žakelj M, Francisci S (eds). Survival of cancer patients in Europe, 1999–2007: the EUROCARE-5 Study. *Eur J Cancer* 2015; **51**(15): 2099–268.

56 www.eurocare.it/Publications/tabid/61/Default.aspx (accessed April 17, 2019).

57 Stellarova-Foucher E, O'Callaghan M, Ferlay J, *et al*. The European Cancer Observatory: a new data resource. *Eur J Cancer* 2015; **51**(9): 1131–43.

58 European Cancer Observatory (ECO). https://ecis.jrc.ec.europea.eu (accessed April 22, 2019).

59 Doherty A, Jackson D, Hammerla N, *et al*. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank Study. *PLoS One* 2017; **12**(2): e0169649.

60 Richardson JL, Stephens S, Thomas SHL, *et al*. An international study of the ability and cost-effectiveness of advertising methods to facilitate study participant self-enrolment into a pilot pharmacovigilance study during early pregnancy. *JMIR Public Health Surveill* 2016; **2**(1): e13.

Part IIId

Choosing a Data Source

17

Choosing among the Available Data Sources for Pharmacoepidemiology Research

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

As discussed in previous chapters, pharmacoepidemiologic studies apply the techniques of epidemiology to the content area of clinical pharmacology. Between 500 and 3000 individuals are usually studied prior to drug marketing. Most postmarketing pharmacoepidemiologic studies need to include at least 10 000 subjects, or draw from an equivalent population for a case–control study, in order to contribute sufficient new information to be worth their cost and effort (see Chapter 4). This large sample size raises logistical challenges. Chapters 10–16 presented many of the different data collection approaches and data resources that have been developed to perform pharmacoepidemiologic studies efficiently, meeting the need for these very large sample sizes. This chapter is intended to synthesize this material, to assist the reader in choosing among the available approaches.

Choosing among the Available Approaches to Pharmacoepidemiologic Studies

Once one has decided to perform a pharmacoepidemiologic study, one needs to decide which of the data collection approaches or data resources described in the earlier chapters of this book should be used. Although to some degree the choice may too often be based upon a researcher's familiarity with given data resources and/or the investigators who have been using them, it is very important to tailor the choice of pharmacoepidemiologic resource to the question to be addressed. One frequently may want to use more than one data collection strategy or resource, in parallel or in combination. If no single resource is optimal for addressing a question, it can be useful to use a number of approaches that complement each other. Indeed, this is probably the preferable approach for addressing important questions. Regardless, investigators are often left with a difficult and complex choice.

In order to explain how to choose among the available pharmacoepidemiologic data resources, it is useful to synthesize the information from the previous chapters on the relative strengths and weaknesses of each of the available pharmacoepidemiologic approaches, examining the comparative characteristics of each (see Table 17.1). One can then examine the characteristics of the research question at hand, in order to choose the pharmacoepidemiologic

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Table 17.1 Comparative characteristics of pharmacoepidemiologic data resources.

Pharmacoepidemiologic approach	Relative size	Relative cost	Relative speed	Representativeness	Population based	Cohort studies possible	Case–control studies possible
Spontaneous reporting	++++	+	++++	++	—	—	+ (with external controls)
Health maintenance organizations/health plans	++	+++	+++	+++	++	++++	++++
Commercial insurance databases	++	+++	+++	+++	++	++++	++++
US government claims databases	+++	++	++	variable	++++	++++	++++
UK medical record databases	++	++	+++	+++	+++	++++	++++
In-hospital databases	+	++	+++	++	—	++	++
Canadian provincial databases	++	++	+++	++++	++++	++++	++++
Pharmacy-based medical record linkage systems	++	++	+++	++++	++++	++++	++++
Ad hoc studies							
Case–control surveillance	variable	+++	+	variable	—	—	++++
Prescription Event Monitoring	+++	+++	+	+++	++	++++	+ (nested)
Registries	variable	+++	+	variable	variable	+++	+++
Field studies							
Ad hoc case–control studies	as feasible	+++	+	as desired	as desired	—	++++
Ad hoc cohort studies	as feasible	++++	—	as desired	as desired	++++	++ (nested)
Randomized trials	as feasible	++++	—	—	—	++++	++ (nested)

Spontaneous reporting	+++	++	—	+++	+++	N/A
Health plans	++++	+++	++	—	++	3–15%/year
Commercial insurance databases	++++	+++	++	—	++	about 25%/year
US government claims databases	++++	+++	++	—	++	variable
UK medical record databases	+++	++++	++	—	++	nil
In-hospital databases	++++	+++	++	++++	—	nil
Canadian provincial databases	++++	+++	++	—	++	nil
Pharmacy-based medical record linkage systems	++++	+	+	—	—	nil
Ad hoc studies						
Case–control surveillance	++	++++	+++	—	+	N/A
Prescription Event Monitoring	+++	+++	++	—	+++	variable
Registries	+++	+++	++	+	variable	N/A
Field studies						
Ad hoc case–control studies	++	++++	+++	++	+	N/A
Ad hoc cohort studies	+++	+++	+++	++	++++	variable
Randomized trials	++++	+++	++++	++	++++	N/A

Note: See the chapter text for descriptions of the column headings, and previous chapters for descriptions of the data resources. N/A, not applicable.

approach best suited to addressing that question (see Table 17.2). The assessment and weights provided in this discussion and in the accompanying tables are arbitrary. They are not being represented as a consensus of the pharmacoepidemiologic community, but represent the judgment of this author alone, based on the material presented in earlier chapters of this book. Nevertheless, I think that most would agree with the general principles described, and even many of the relative ratings. My hope is that this synthesis of information, despite some of the arbitrary ratings inherent in it, will make it easier for the reader to synthesize the large amount of information presented in prior chapters.

Note that there are a number of other data sources not discussed here, some of which have been, or in the future may be, of importance to pharmacoepidemiologic research. Examples include the old Boston Collaborative Drug Surveillance data [1], MEMO [2], Pharmetrics® [3], Aetna [4], Humana [5], and many others, many reviewed in prior editions of this book. Given the wonderful proliferation of pharmacoepidemiologic data resources, we are making no attempt to include them all. Instead, we will discuss them in categories of types of data, as we did in the chapters themselves.

Comparative Characteristics of Pharmacoepidemiologic Data Resources

Table 17.1 lists each of the different pharmacoepidemiologic data resources that were described in earlier chapters, along with some of their characteristics.

The *relative size* of the database refers to the population it covers. Only spontaneous reporting systems, US Medicare, some of the pharmacy-based medical record linkage systems, and Prescription Event Monitoring in the UK cover entire countries or large fractions thereof. Of course, population databases differ considerably in size, based on the size of their underlying populations. Aggregations of Medicaid databases are the next largest, with the commercial databases approaching that. The UK electronic health record databases would be next in size, as would the health maintenance organizations (HMOs), depending on how many are included. The Canadian provincial databases again could be equivalently large, depending in part on how many are included in a study. The other data resources are generally smaller. Case–control surveillance, as formerly conducted by the Slone Epidemiology Unit, can cover a variable population, depending on the number of hospitals and metropolitan areas included in the network for a given study. The population base of registry-based case–control studies depends on the registries used for case finding. *Ad hoc* studies can be whatever size the researcher desires and can marshal resources for.

As to *relative cost*, studies that collect new data are most expensive, especially randomized trials and cohort studies, for which sample sizes generally need to be large and follow-up may need to be prolonged. In the case of randomized trials, there are additional logistical complexities. Studies that use existing data are least expensive, although their cost increases when they gather primary medical records for validation. Studies that use existing data resources to identify subjects but then collect new data about those subjects are intermediate in cost.

With regard to the *relative speed* of study completion, studies that collect new data take longer, especially randomized trials and cohort studies. Studies that use existing data are able to answer a question most quickly, although considerable additional time may be needed to obtain primary medical records for validation. Studies that use existing data resources to identify subjects but then collect new data about those subjects are intermediate in speed.

Representativeness refers to how well the subjects in the data resource represent the population at large or a more specific population of

Table 17.2 Characteristics of research questions and their impact on the choice of pharmacoepidemiologic data resources.

Pharmacoepidemiologic approach	Hypothesis generating[a]	Hypothesis strengthening[b]	Hypothesis testing[c]	Study of benefits (versus risk)	Incidence rates desired	Low incidence outcome	Low prevalence exposure
Spontaneous reporting	++++	+	—	—	—	++++	++++
Health plans	++	++++	+++	++	+++	+++	+++
Commercial insurance databases	++	++++	+++	++	+++	+++	+++
US government claims databases	++	++++	+++	++	+++	++++	++++
UK medical record databases	++	++++	+++	++	++++	+++	+++
In-hospital databases	+	++++	+++	++	+++	+	+
Canadian provincial databases	++	++++	+++	++	+++	+++	+++
Pharmacy-based medical record linkage systems	+	++	++	++	+++	+++	+++
Ad hoc studies							
Case–control surveillance	+++	+++	+++	+++	—	++++	+
Prescription Event Monitoring	++	++	+++	+++	+++	+++	+++
Registries	+	+++	+++	+++	+++	+++	+++
Field studies							
Ad hoc case–control studies	+	++	+++	+++	+	++++	+
Ad hoc cohort studies	+	++	+++	+++	++++	++	+++
Randomized trials	+	+	++++	++++	++++	+	++++

(*Continued*)

Table 17.2 (Continued)

Pharmacoepidemiologic approach	Important confounders	Drug use inpatient (versus outpatient)	Outcome does not result in hospitalization	Outcome does not result in medical attention	Outcome a delayed effect	Exposure to a new drug	Urgent question
Spontaneous reporting	—	+++	++++	+	+	++++	++++
Health plans	+++	—	+++	—	+	++	+++
Commercial insurance databases	++	—	+++	—	+	+++	+++
US government claims databases	++	—	+++	—	+ to +++	++	++
UK medical record databases	+++	—	+++	—	+++	+++	+++
In-hospital databases	++	++++	—	—	—	+++	+++
Canadian provincial databases	++	—	+++	—	+++	++	+++
Pharmacy-based medical record linkage systems	+	—	—	—	++	+++	+++
Ad hoc studies							
Case–control surveillance	+++	+	—	—	++	+	+
Prescription Event Monitoring	++	+	++++	+	+	++++	+
Registries	++	++	+	++	++	+++	+
Field studies							
Ad hoc case–control studies	+++	++++	++	—	++	+	+
Ad hoc cohort studies	+++	+++	++++	+++	+	++++	+
Randomized trials	++++	+++	++++	++++	+	++++	+

Notes: See the text of the chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.
[a] Hypothesis-generating studies are designed to raise new questions about possible unexpected drug effects, whether adverse or beneficial.
[b] Hypothesis-strengthening studies are designed to provide support for, although not definitive evidence for, existing hypotheses.
[c] Hypothesis-testing studies are designed to evaluate in detail hypotheses raised elsewhere.

interest. US Medicare, Prescription Event Monitoring in the UK, the provincial health databases in Canada, and the pharmacy-based medical record linkage systems include entire countries, provinces, or states, and so are typical populations. Spontaneous reporting systems are drawn from entire populations, but of course the selective nature of their reporting could lead to less certain representativeness. Medicaid programs are limited to the disadvantaged, and so include a population that is least representative of a general population. Analogously, randomized trials include populations limited by the various selection criteria plus their willingness to volunteer for the study. The Clinical Practice Research Datalink® (CPRD®) and The Health Improvement Network® (THIN®) use a nonrandom large subset of the total UK population, and so may be representative of the overall UK population. Health plans and commercial databases are closer to representative populations than a Medicaid population would be, although they include a largely working population and, so, include few patients of low socioeconomic status and fewer than normal elderly. Some of the remaining data collection approaches or resources are characterized in Table 17.1 as "variable," meaning their representativeness depends on which hospitals are recruited into the study. *Ad hoc* studies are listed in Table 17.1 "as desired," because they can be designed to be representative or not, as the investigator wishes.

Whether a database is *population based* refers to whether there is an identifiable population (which is not necessarily based on geography), all of whose medical care would be included in that database, regardless of the provider. This allows one to measure incidence rates of diseases, as well as being more certain that one knows of all the medical care that any given patient receives. As an example, assuming little or no out-of-plan care, the Kaiser programs are population based. One can use Kaiser data, therefore, to study medical care received in and out of the hospital, as well as diseases that may result in repeat hospitalizations. For example, one could study the impact of the treatment initially received for venous thromboembolism on the risk of subsequent disease recurrence. In contrast, hospital-based case–control studies conducted outside a closed network like Kaiser are not population based: they include only the specific hospitals that belong to the system and do not capture all healthcare services a patient may receive. Thus, a patient diagnosed with and treated for venous thromboembolism in a participating hospital could be readmitted to a different, nonparticipating hospital if the disease recurred. This recurrence would not be detected in a study using such a system. The data resources that are population based are those that use data from organized healthcare delivery or payment systems. Registry-based and *ad hoc* case–control studies can occasionally be conducted as population-based studies, if all cases in a defined geographic area are recruited into the study [6], but this is unusual (see also Chapters 3 and 16).

Whether cohort studies are possible within a particular data resource would depend upon whether individuals can be identified by whether or not they were exposed to a drug of interest. This would be true in any of the population-based systems, as well as any of the systems designed to perform cohort studies.

Whether case–control studies are possible within a given data resource depends upon whether patients can be identified by whether or not they suffered from a disease of interest. This would be true in any of the population-based systems. Data from spontaneous reporting systems can be used for case finding for case–control studies, although this has been done infrequently [7].

The *validity of the exposure data* is most certain in hospital-based settings, where one can be reasonably certain of both the identity of a drug and that the patient actually ingested it. Exposure data in spontaneous reporting systems come mostly from healthcare providers

and so are probably valid. However, one cannot be certain of patient adherence in spontaneous reporting data. Exposure data from claims data and from pharmacy-based medical record linkage systems are unbiased data recorded by pharmacies, often for billing purposes, a process that is closely audited as it impacts reimbursement. These data are likely to be accurate with regard to medication possession, although, again, one cannot assure adherence. Refill adherence, though, has been found to correlate closely with adherence measured using microchips embedded in medication bottles (see Chapter 38). However, there are drugs that may fall beneath a patient's deductibles or co-payments, or not be on formularies, so dispensed by the pharmacy but paid for in cash. In claims databases, these scenarios may result in misclassification of true medication exposure, as the patient would falsely appear unexposed. Also, since drug benefits vary depending upon the plan, pharmacy files may not capture all prescribed drugs if beneficiaries reach the drug benefit limit or pay for the prescription out of pocket. In the UK medical record systems, drugs prescribed by physicians other than the general practitioner could be missed, although continued prescribing by the general practitioner would be detected. *Ad hoc* case–control studies generally rely on patient histories for exposure data. These may be very inaccurate, as patients often do not recall correctly the medications they are taking [8]. However, this would be expected to vary, depending upon the condition studied, type of drug taken, questioning technique used, and so on [8–16] (see Chapter 37).

The *validity of the outcome data* is also most certain in hospital-based settings, in which the patient is subjected to intensive medical surveillance (see Chapter 14). It is least certain in outpatient data from organized systems of medical care. There are, however, methods of improving the accuracy of these data, such as using drugs, laboratory data, and procedures as markers of the disease and obtaining primary medical records (see Chapter 37). The outcome data from automated databases are listed as variable, therefore, depending upon exactly which data are being used and how. The UK medical record systems analyze the actual medical record, rather than claims, and can access additional questionnaire data from the general practitioner as well. Thus, their outcome data may be more accurate.

Control of confounding refers to the ability to control for confounding variables. As discussed in Chapter 3, randomization is the most convincing way of controlling for unknown, unmeasured, or unmeasurable confounding variables. Approaches that collect sufficient information to control for known and measurable variables are next most effective. These include health plans, the UK medical record systems, case–control surveillance, *ad hoc* case–control studies, and *ad hoc* cohort studies. Users of health databases in Canada, commercial databases, and Medicaid (sometimes) can obtain primary medical records, but not all information necessary is always available in those records. They generally are unable to contact patients directly to obtain supplementary information that might not be in a medical record. Finally, spontaneous reporting systems do not provide enough systematically collected information for control of confounding.

Relatively few of the data systems have data on *inpatient drug use*. The exceptions include spontaneous reporting systems, in-hospital databases (see Chapter 14), and some *ad hoc* studies if designed to collect such.

Only a few of the data resources have sufficient *data on outpatient diagnoses* available without special effort to be able to study them as outcome variables. *Ad hoc* studies can be designed to be able to collect such information. In the case of *ad hoc* randomized clinical trials, this data collection effort could even include tailored laboratory and physical examination measurements. In some of the resources, the

outpatient outcome data are collected observationally, but directly via the physician, and so are more likely to be accurate. Included are spontaneous reporting systems, the UK medical record systems, HMOs, Prescription Event Monitoring, and some *ad hoc* cohort studies. Other outpatient data come via physician claims for medical care, including Medicaid databases, commercial databases, and the provincial health databases in Canada. Finally, other data resources can access outpatient diagnoses only via the patient, so they are less likely to be complete; although the diagnosis can often be validated using medical records, it generally needs to be identified by the patient. These include most *ad hoc* case–control studies.

The degree of *loss to follow-up* differs substantially among the different resources. They are specified in Table 17.1.

Characteristics of Research Questions and Their Impact on the Choice of Pharmacoepidemiologic Data Resources

Once one is familiar with the characteristics of the pharmacoepidemiologic resources available, one must then examine more closely the research question, to determine which resources can best be used to answer it (see Table 17.2).

Pharmacoepidemiologic studies can be undertaken to generate hypotheses about drug effects, to strengthen hypotheses, and/or to test *a priori* hypotheses about drug effects. *Hypothesis-generating studies* are studies designed to raise new questions about possible unexpected drug effects, whether adverse or beneficial. Virtually all studies can and do raise such questions, through incidental findings in studies performed for other reasons. In addition, virtually any case–control study could be used, in principle, to screen for possible drug causes of a disease under study, and virtually any cohort study could be used to screen for unexpected outcomes from a drug exposure under study. In practice, however, the only settings in which this has been attempted systematically have been health plans, case–control surveillance, Prescription Event Monitoring, and Medicaid databases. To date, the most productive source of new hypotheses about drug effects has been spontaneous reporting. However, this is the goal of Sentinel, a Congressionally mandated data system of over 100 million US lives, initially built primarily for hypothesis strengthening as "Mini-Sentinel," although now being used for hypothesis generation as well, in addition to the traditional approach of using such data for hypothesis testing (see Chapter 25). In the future, new approaches using the internet (e.g., health websites with consumer posting boards and other social media) could potentially be used for hypothesis generation of events, including those not coming to medical attention.

Hypothesis-strengthening studies are designed to provide support for, although not definitive evidence for, existing hypotheses. The objective of these studies is to provide sufficient support for, or evidence against, a hypothesis to permit a decision about whether a subsequent, more definitive study should be undertaken. As such, hypothesis-strengthening studies need to be conducted rapidly and inexpensively. They can include crude analyses conducted using almost any dataset, evaluating a hypothesis which arose elsewhere. Because not all potentially confounding variables would be controlled, the findings could not be considered definitive. Examples would be the modular studies conducted within Sentinel (see Chapter 25). Alternatively, hypothesis-strengthening studies can be more detailed, controlling for confounding, conducted using the same data resource that raised the hypothesis. In this case, because the study is not specifically undertaken to test an *a priori* hypothesis, the hypothesis-testing type of study can only serve to strengthen, not test, the hypothesis. Spontaneous reporting systems are useful for raising hypotheses, but are

not very useful for providing additional support for those hypotheses. Conversely, randomized trials can certainly strengthen hypotheses, but are generally too costly and logistically too complex to be used for this purpose. (*Post-hoc* analyses of randomized trials can obviously be reanalyzed, for the purposes of generating or strengthening hypotheses, but then they are really being analyzed as cohort studies.) Of the remaining approaches, those that can quickly access, in computerized form, both exposure data and outcome data are most useful. Those that can rapidly access only one of these data types, only exposure or only outcome data, are next most useful, while those that need to gather both data types are least useful, because of the time and expense that would be entailed.

Hypothesis-testing studies are designed to evaluate in detail hypotheses raised elsewhere. Such studies must be able to have simultaneous comparison groups and must be able to control for most known potential confounding variables. For these reasons, spontaneous reporting systems cannot be used for this purpose, as they cannot be used to conduct studies with simultaneous controls (with rare exceptions, see [2]). The most powerful approach, of course, is a randomized clinical trial, as it is the only way to control for unknown or unmeasurable confounding variables. Instrumental variable analyses can approximate a randomized clinical trial, but only in the circumstances, to date limited, that all the underlying assumptions are met. (On the other hand, studies of dose response, duration response, drug–drug interactions, determinants of response, etc. are more readily done in nonrandomized than randomized studies; see Chapter 3.) Techniques which allow access to patients and their medical records are the next most powerful, as one can gather information on potential confounders that might only be reliably obtained from one of those sources or the other. Techniques which allow access to primary records but not the patient are next most useful.

The research implications of questions about the *beneficial effects* of drugs are different, depending upon whether the beneficial effects of interest are expected or unexpected. Studies of *unexpected beneficial effects* are exactly analogous to studies of unexpected adverse effects, in terms of their implications for one's choice of approach; in both situations one is studying side effects. Studies of *expected beneficial effects*, or drug efficacy, raise the special methodologic problem of confounding by the indication: patients who receive a drug are different from those who do not in a way which usually is related to the outcome under investigation in the study. This issue is discussed in detail in Chapter 33. As described there, it *is* sometimes possible to address these questions using nonexperimental study designs. Generally, however, the randomized clinical trial is far preferable, when feasible.

In order to address questions about the *incidence of a disease* in those exposed to a drug, one must be able to quantify how many people received the drug. This information can be obtained using any resource that can perform a cohort study. Techniques that need to gather the outcome data *de novo* may miss some of the outcomes if there is incomplete participation and/or reporting of outcomes, such as with Prescription Event Monitoring, *ad hoc* cohort studies, and outpatient pharmacy-based cohort studies. On the other hand, *ad hoc* data collection is the only way of systematically collecting information about outcomes that need not come to medical attention (see below). The only approaches that are free from either of these problems are hospital-based approaches. Registry-based case–control studies and *ad hoc* case–control studies can occasionally be used to estimate incidence rates, if one obtains a complete collection of cases from a defined geographic area. The other approaches listed cannot be used to calculate incidence rates.

To address a question about a *low incidence outcome*, one needs to study a large population

(see Chapter 4). This can best be done using spontaneous reporting, US Medicare, Prescription Event Monitoring, or the pharmacy-based medical record linkage systems, which can or do cover entire countries. Alternatively, one could use commercial databases, health plans, or aggregates of Medicaid databases, which cover a large proportion of the US, or the medical record systems in the UK. Canadian provincial databases can also be fairly large, and one can perform a study in multiple such databases. *Ad hoc* cohort studies could potentially be expanded to cover equivalent populations. Case–control studies, either *ad hoc* studies, studies using registries, or studies using case–control surveillance, can also be expanded to cover large populations, although not as large as the previously mentioned approaches. Because case–control studies recruit study subjects on the basis of the patients suffering from a disease, they are more efficient than attempting to perform such studies using analogous cohort studies. Finally, randomized trials could, in principle, be expanded to achieve very large sample sizes, especially large simple trials (see Chapter 32), but this can be extremely difficult and costly.

To address a question about a *low prevalence exposure*, one also needs to study a large population (see Chapter 4). Again, this can best be done using spontaneous reporting, US Medicare, the pharmacy-based medical record linkage systems, or Prescription Event Monitoring, which cover entire countries. Alternatively, one could use commercial databases, large health plans, or aggregates of Medicaid databases, which cover a large proportion of the US, or the medical record databases in the UK. *Ad hoc* cohort studies could also be used to recruit exposed patients from a large population. Analogously, randomized trials, which specify exposure, could assure an adequate number of exposed individuals. Case–control studies, either *ad hoc* studies, studies using registries, or studies using case–control surveillance, could theoretically be expanded to cover a large enough population, but this would be difficult and expensive.

When there are *important confounders* that need to be taken into account in order to answer the question at hand, then one needs to be certain that sufficient and accurate information is available on those confounders. Spontaneous reporting systems cannot be used for this purpose. The most powerful approach is a randomized trial, as it is the most convincing way to control for unknown or unmeasurable confounding variables. Techniques which allow access to patients and their medical records are the next most powerful, as one can gather information on potential confounders that might only be reliably obtained from one of those sources or the other. Techniques which allow access to primary records but not the patient are the next most useful.

If the research question involves *inpatient* drug use, then the data resource must obviously be capable of collecting data on inpatient drug exposures. The number of approaches that have this capability are limited, and include spontaneous reporting systems and inpatient database systems. *Ad hoc* studies could also, of course, be designed to collect such information in the hospital.

When the *outcome under study does not result in hospitalization, but does result in medical attention*, the best approaches are randomized trials and *ad hoc* studies, which can be specifically designed to be sure this information can be collected. Prescription Event Monitoring and the UK medical record systems, which collect their data from general practitioners, are excellent sources of data for this type of question. Reports of such outcomes are likely to come to spontaneous reporting systems as well. Medicaid databases and commercial databases can also be used, as they include outpatient data, although one must be cautious about the validity of the diagnosis information in outpatient claims. Canadian provincial databases are

similar, as are health plans. Finally, registry-based case–control studies could theoretically be performed, if they included outpatient cases of the disease under study.

When the *outcome under study does not result in medical attention at all*, the approaches available are much more limited. Only randomized trials and prospective cohort studies can be specifically designed to be certain this information is collected. Finally, occasionally one could collect information on such an outcome in a spontaneous reporting system, if the report came from a patient or from a healthcare provider who became aware of the problem while the patient was visiting for medical care for some other problem. In the future, as already noted, new approaches using the internet (e.g., health websites with consumer posting boards) could potentially be used for hypothesis generation of events not coming to medical attention.

When the *outcome under study is a delayed drug effect*, then one obviously needs approaches capable of tracking individuals over a long period of time. The best approach for this are some of the provincial health databases in Canada. Drug data are available in some of these for more than 25 years, and there is little turnover in the population covered. Thus, this is an ideal system within which to perform such long-term studies. Some health plans have even longer follow-up time available. However, as health plans they suffer from substantial turnover, albeit more modest after the first few years of enrollment. Commercial databases are similar. Any of the methods of conducting case–control studies can address such questions, although one would have to be especially careful about the validity of exposure information collected many years after the exposure. Medicaid databases have been available since 1973. However, the large turnover in Medicaid programs, due to changes in eligibility with changes in family and employment status, makes studies of long-term drug effects problematic. Similarly, one could conceivably perform studies of long-term drug effects using Prescription Event Monitoring, the pharmacy-based medical record linkage systems, *ad hoc* cohort studies, or randomized clinical trials, but these approaches are not as well suited to this type of question as the previously discussed techniques. Theoretically, one also could identify long-term drug effects in a spontaneous reporting system. This is improbably, however, as a physician is unlikely to link a current medical event with a drug exposure long ago.

When *the exposure under study is a new drug*, then one is, of course, limited to data sources that collect data on recent exposures, and preferably those that can collect a significant number of such exposures quickly. *Ad hoc* cohort studies or a randomized clinical trial are ideal for this, as they recruit patients into the study on the basis of their exposure. Spontaneous reporting is similarly a good approach, as new drugs are automatically and immediately covered, and in fact reports are much more common in the first three years after a drug is marketed. The major databases are next most useful, especially the commercial ones, as their large population base will allow one to accumulate a sufficient number of exposed individuals rapidly, so one can perform a study sooner. In some cases, there is a delay until the drug is available on the program's formulary; however, that especially can be an issue with HMOs. The US government claims databases (Medicare and Medicaid) have a delay in availability of their data, which makes them less useful for the newest drugs. *Ad hoc* case–control studies, by whatever approach, must wait until sufficient drug exposure has occurred that it can affect the outcome variable being studied.

Finally, if *one needs an answer to a question urgently*, potentially the fastest approach, if the needed data are included, is a spontaneous reporting system; drugs are included in these systems immediately, and an extremely large population base is covered. Of course, one cannot rely on any adverse reaction being detected

in a spontaneous reporting system. The computerized databases are also useful for these purposes, depending on the speed with which the exposures accumulate in them; of course, if the drug is not on the formulary in question, it cannot be studied. Modular analyses in Sentinel were designed for exactly this purpose (see Chapter 25). The remaining approaches are of limited use, as they take too long to address a question. One exception to this is Prescription Event Monitoring, if the drug in question happens to have been a subject of one of its studies. The other, and more likely, exception is case–control surveillance, if the disease under study is available in adequate numbers in its database, either because it was the topic of a prior study or because there was a sufficient number of individuals with the disease collected to be included in control groups for prior studies.

Examples

As an example, one might want to explore whether nonsteroidal anti-inflammatory drugs (NSAIDs) cause upper gastrointestinal bleeding and, if so, how often. One could examine the manufacturer's premarketing data from clinical trials, but the number of patients included is not likely to be large enough to study clinical bleeding, and the setting is very artificial. Alternatively, one could examine premarketing studies using more sensitive outcome measures, such as endoscopy. However, these are even more artificial. Instead, one could use any of the databases to address the question quickly, as they have data on drug exposures that preceded the hospital admission. Some databases could only be used to investigate gastrointestinal bleeding resulting in hospitalization (e.g., Kaiser Permanente, except via chart review). Others could be used to explore inpatient or outpatient bleeding (e.g., Medicare, Medicaid, Canadian provincial databases). Because of confounding by cigarette smoking, alcohol, and so on, which would not be well measured in these databases, one also might want to address this question using case–control or cohort studies, whether conducted *ad hoc* or using any of the special approaches available, for example case–control surveillance or Prescription Event Monitoring. If one wanted to be able to calculate incidence rates, one would need to restrict these studies to cohort studies, rather than case–control studies. One would be unlikely to be able to use registries, as there are no registries, known to this author at least, which record patients with upper gastrointestinal bleeding. One would not be able to perform analyses of secular trends, as upper gastrointestinal bleeding would not appear in vital statistics data, except as a cause of hospitalization or death. Studying death from upper gastrointestinal bleeding is problematic, as it is a disease from which patients usually do not die. Rather than studying determinants of upper gastrointestinal bleeding, one would really be studying determinants of complications from upper gastrointestinal bleeding, diseases for which upper gastrointestinal bleeding is a complication, or determinants of physicians' decisions to withhold supportive transfusion therapy from patients with upper gastrointestinal bleeding, for example age, terminal illnesses, and so on.

Alternatively, one might want to address a comparable question about nausea and vomiting caused by NSAIDs. Although this question is very similar, one's options in addressing it would be much more limited, as nausea and vomiting often do not come to medical attention. Other than a randomized clinical trial, for a drug that is largely used on an outpatient basis one is limited to systems which request information from patients, or *ad hoc* cohort studies.

As another example, one might want to follow up on a signal generated by the spontaneous reporting system, designing a study to investigate whether a drug which has been on the market for, say, five years is a cause of a relatively rare condition, such as allergic hypersensitivity

reactions. Because of the infrequency of the disease, one would need to draw on a very large population. The best alternatives would be Medicare or Medicaid databases, health plans, commercial databases, case–control studies, or Prescription Event Monitoring. To expedite this hypothesis-testing study and limit costs, it would be desirable if it could be performed using existing data. Prescription Event Monitoring and case–control surveillance would be excellent ways of addressing this, but only if the drug or disease in question, respectively, had been the subject of a prior study. Other methods of conducting case–control studies require gathering exposure data *de novo*.

As a last example, one might want to follow up on a signal generated by a spontaneous reporting system, designing a study to investigate whether a drug which has been on the market for, say, three years is a cause of an extremely rare but serious illness, such as aplastic anemia. One's considerations would be similar to those just described, but even Medicare or Medicaid databases would not be sufficiently large to include enough cases, given the delay in the availability of their data. One would have to gather data *de novo*. Assuming the drug in question is used mostly by outpatients, one could consider using Prescription Event Monitoring or a case–control study.

Conclusion

Once one has decided to perform a pharmacoepidemiologic study, one needs to decide which of the resources described in the earlier chapters of this book should be used. By considering the characteristics of the pharmacoepidemiologic resources available as well as the characteristics of the question to be addressed, one should be able to choose those resources that are best suited to addressing the question at hand.

References

1 Lawson DH, Beard K. *Intensive hospital-based cohort studies*. In: Strom BL, ed. Pharmacoepidemiology, 2nd edn. Chichester: John Wiley & Sons, 1994, pp. 157–70.

2 Wei L, Parkinson J, MacDonald TM. The Tayside Medicines Monitoring Unit (MEMO). In: Strom BL, ed. *Pharmacoepidemiology*, 4th edn. Chichester: John Wiley & Sons, 2005, pp. 323–36.

3 Prescott JD, Factor S, Pill M, Levi GW. Descriptive analysis of the direct medical costs of multiple sclerosis in 2004 using administrative claims in a large nationwide database. *J Manag Care Pharm* 2007; **13**: 44–52. http://www.amcp.org/data/jmcp/44-52.pdf

4 Maddison P, Kiely P, Kirkham B, *et al*. Leflunomide in rheumatoid arthritis: recommendations through a process of consensus. *Rheumatology (Oxford)* 2005; **44**: 280–6.

5 Shafazand S, Yang Y, Amore E, O'Neal W, Brixner D. A retrospective, observational cohort analysis of a nationwide database to compare heart failure prescriptions and related health care utilization before and after publication of updated treatment guidelines in the United States. *Clin Ther* 2010; **32**: 1642–50.

6 Risks of agranulocytosis and aplastic anemia. A first report of their relation to drug use with special reference to analgesics. The International Agranulocytosis and Aplastic Anemia Study. *JAMA* 1986; **256**: 1749–57.

7 Strom BL, West SL, Sim E, Carson JL. Epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693–9.

8 Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health* 1967; **57**: 2071–5.

9 Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**: 273–5.

10 Stolley PD, Tonascia JA, Sartwell PE, *et al*. Agreement rates between oral contraceptive users and prescribers in relation to drug use histories. *Am J Epidemiol* 1978; **107**: 226–35.

11 Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114–22.

12 Rosenberg MJ, Layde PM, Ory HW, Strauss LT, Rooks JB, Rubin GL. Agreement between women's histories of oral contraceptive use and physician records. *Int J Epidemiol* 1983; **12**: 84–7.

13 Schwarz A, Faber U, Borner K, Keller F, Offermann G, Molzahn M. Reliability of drug history in analgesic users. *Lancet* 1984; **2**: 1163–4.

14 Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. *Contraception* 1986; **33**: 127–39.

15 Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.

16 Persson I, Bergkvist L, Adami HO. Reliability of women's histories of climacteric oestrogen treatment assessed by prescription forms. *Int J Epidemiol* 1987; **16**: 222–8.

Part IV

Selected Applications of Pharmacoepidemiology

18

Studies of Drug Utilization

Björn Wettermark[1], Vera Vlahović-Palčevski[2], David Lee[3], and Ulf Bergman[1]

[1] *Division of Clinical Pharmacology, Department of Laboratory Medicine, Centre for Pharmacoepidemiology, Karolinska Institute, Karolinska University Hospital-Huddinge, Stockholm, Sweden*
[2] *Department of Clinical Pharmacology, University Hospital Rijeka, Department of Pharmacology, University of Rijeka Medical Faculty and Department of Basic Medical Sciences, Faculty of Health Studies, Rijeka, Croatia*
[3] *Pharmaceuticals and Health Technologies Group, Management Sciences for Health, Arlington, VA, USA*

Historical Background

Interest in drug utilization studies began on both sides of the Atlantic in the early 1960s. There was recognition of the virtual explosion in the marketing of new drugs, the wide variations in the patterns of drug prescribing and consumption, growing concern about delayed adverse effects, and increasing concern about drug expenditure, as reflected in the increase in both the monetary sales and the volume of drug prescriptions [1,2]. However, the development of pharmacoepidemiologic methods can be characterized by two different lines of work (drug utilization studies as performed in Europe versus pharmacoepidemiology as performed in the US), approaching each other from opposite directions, strongly influenced by variations in availability and accessibility of data sources.

Drug utilization studies at the national and international levels have been more developed in Europe, and pioneered by the Nordic countries, Scotland, the Czech Republic, and Northern Ireland. Under the auspices of the World Health Organization (WHO) Regional Office for Europe, a Drug Utilization Research Group was established in the 1970s to stimulate interest in studies comparing drug utilization between countries using a common methodology [1]. Factors that contributed greatly to this line of development, primarily in the countries of Northern Europe, have been the relatively small size of the populations involved, the limited number of pharmaceutical products on the market (2000–3000 in Norway and Sweden), and the availability of centralized statistics on wholesaler sales or dispensed prescriptions [1]. The early drug utilization studies were conducted during a time when data were not computerized and when there was no uniform classification system for medicines. However, with the growth of registers and computer technology, the size of the population is less of an issue and studies can be conducted on large populations. During the last decades, national databases with patient-level data on prescription medicines have been established in most European countries, but there is still a lack of comprehensive databases on medicines used in inpatient care [3,4].

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Drug utilization studies in Europe were originally predominantly quantitative, describing and comparing patterns of utilization of specific groups of drugs according to geographic regions and time. For example, international comparative studies have documented wide variations in the utilization of antidiabetic [1,5], psychotropic [6], nonsteroidal anti-inflammatory drugs (NSAIDs) [7,8], antihypertensive [1,6], antibiotic [9], and lipid-lowering drugs [10] in European and other countries. Longitudinal studies on the utilization of antidiabetic and antihypertensive drugs in some of these countries indicate that the differences cannot be explained only by differences in the prevalence of disease [11,12], and studies on, for instance, lipid-lowering agents have shown that the large increases in statin utilization were not associated with a subsequent decrease in coronary heart disease (CHD) mortality [13]. This illustrates that drug utilization patterns are complex and dependent on a range of factors. Furthermore, there is substantial room for improvement in the quality of medicines use; that is, how patients use medicines.

National studies have also revealed striking variations in drug utilization between regions and communities within the same country [1,6]. However, most of these studies have been descriptive and only a few of them have addressed the relation between variations in drug sales and treatment outcomes [14]. Since many of the studies have an ecologic design, examining associations between exposure and outcome in populations rather than individuals, they cannot directly be interpreted as associations at the level of the individual.

In Canada and the US, drug utilization research developed on a smaller scale, primarily at institutional or local health program levels. Factors that have hindered studies at a national level were originally the size of the population, the number of pharmaceutical products on the market (20 000–30 000), and the lack of an all-encompassing pharmaceutical data collection system [15]. Data on drug use are more readily available from health plans, health delivery institutions, and public healthcare programs. For example, early studies of physician prescribing showed that prescribing patterns varied greatly among physicians, according to their place and type of practice and the community in which they prescribed [16]. North American drug utilization research placed greater emphasis on studying the quality of physician prescribing practices, in particular with respect to antibiotics, in both hospital and outpatient settings [17–19]. This was followed by studies that targeted medications for cardiovascular diseases [20–23]. Studies describing national patterns of drug utilization and expenditures in the US are scarce [15], while those addressing the use of various types of medications, including herbal and other natural products, in adults and children are performed more often [24–27].

Drug utilization research has also developed in Latin America, Australia, Asia, and Africa. In Latin America, a drug utilization research group (the Latin American Group for Drug Utilization, DURG-LA) was founded in 1991 [28]. Over the years Latin American drug utilization researchers have conducted a wide range of studies on rational use of medicines, often using primary data collection in people's homes, in pharmacies, or in health facilities. Secondary data on drug utilization have traditionally been fragmented and difficult to access in Latin America [29]. There has been a rapid development of drug utilization research in low- and middle-income countries. In the early 1990s, the WHO and the International Network for the Rational Use of Drugs (INRUD) published a simple sampling method and a standard set of indicators to describe core aspects of prescribing and dispensing [30]. The first International Conference on Improving the Use of Medicines (ICIUM), held in Chiang Mai, Thailand, in 1997, systematically reviewed interventions to promote rational drug use in developing countries [30]. Substantial problems of irrational use of

medicines were identified and some key areas were highlighted for future research, such as interventions to improve the use of antibiotics and antimalarial drugs, methods to assess the impact of Drugs and Therapeutic Committees, and the impact of financial incentives on drug utilization patterns. During the last decade, we have witnessed a rapid growth of large prescription databases in Asia [31]. It is likely that these will further contribute to the globalization of drug utilization research.

Definitions

Drug utilization research can be defined as "an eclectic collection of descriptive and analytical methods for the quantification, the understanding and the evaluation of the processes of prescribing, dispensing and consumption of medicines and for the testing of interventions to enhance the quality of these processes" [32]. This definition includes both quantitative and qualitative research methods.

The WHO defined drug utilization as the "marketing, distribution, prescription and use of drugs in a society, with special emphasis on the resulting medical, social, and economic consequences" [33]. Some authors have suggested that the development of drugs relative to health priorities should also be included in studies of drug utilization [34]. This broad definition differs from the narrower one that appeared in the North American literature, "the prescribing, dispensing and ingesting of drugs" [35,36].

In the US, drug utilization review (DUR), or drug use evaluation (DUE), refers to an authorized, structured, ongoing review of prescribing, dispensing, and use of medication (see Chapter 19). It involves a comprehensive review of patients' prescription and medication data before, during, and/or after dispensing to ensure appropriate medication decision making and positive patient outcomes. As such, DUR is a quality assurance measure [37].

In all of these definitions, recognition is granted, explicitly or implicitly, of the nonclinical (e.g., socio-anthropological, behavioral, and economic) factors influencing drug utilization. Studies of the process of drug utilization focus on the factors influencing and events involved in the prescribing, dispensing, administration, and taking of medication. However, the broader definitions of the WHO, the Academy of Managed Care Pharmacy (AMCP), and the European Drug Utilization Group go beyond the "process" of drug utilization, which is the movement of drugs along the therapeutic drug chain, to include consideration of the various outcomes, such as use of drugs of doubtful or no clinical efficacy, and the quality of drug use [38]; that is, the degree to which it adheres to established norms. According to these definitions, studies of drug utilization include not only studies of the medical and nonmedical factors influencing drug utilization, but also the effects of drug utilization at all levels, from the individual patient to the society. Studies of how drug utilization relates to the effects of drug use, beneficial or adverse, are usually labeled analytic pharmacoepidemiologic research. These two aspects of the study of drug utilization have developed along parallel lines, but may now be regarded as interrelated and part of a continuum of interests and methods.

As stated by Lunde and Baksaas [39], the general objectives of drug utilization studies are:

> problem identification and problem analysis in relation to importance, causes, and consequences; establishment of a weighted basis for decisions on problem solution; assessment of the effects of the action taken. These objectives are relevant to problems and decision making throughout the drug and health chain. The approaches may vary according to the purpose and the needs of the users. Those include the health authorities, the drug manufacturers, the academic and

clinical health professionals, social scientists, and economists as well as the media and the consumers.

Since many drug utilization studies have a strong focus on health policy, the discipline may be seen as the bridge between pharmacoepidemiology and health services research. It is also closely connected to clinical pharmacology, with the principal aim of drug utilization research being to facilitate the safe and effective use of medicines in populations [40].

This chapter focuses on the current status of descriptive epidemiologic approaches to the study of the processes of drug utilization and analytic studies on factors associated with drug utilization patterns. The epidemiologic approaches to the study of the beneficial and harmful effects of drug utilization are covered elsewhere in this book.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

In order for a drug to be marketed, it must be shown that it can effectively modify the natural course of disease or alleviate symptoms when used appropriately, for the right patient, with the right disease, in the proper dosage and intervals, and for the appropriate length of time. Used inappropriately, however, drugs often fail to live up to their potential, with consequent morbidity and mortality and waste of resources.

Drug utilization research describes the extent and pattern, quality, determinants, and outcomes of drug exposure. Pattern of use covers the extent, profiles, and trends in drug use and costs over time. It gives answers and helps understand how drugs are used in terms of incidence, prevalence, and trends over time. Quality of use is determined using audits to compare actual use to national prescription guidelines or local drug formularies. Indicators of quality of drug use may include the choice of drug (adherence with the guideline or formulary), drug cost (compliance with budgetary recommendations), drug dosage (awareness of interindividual variations in dose requirements and age dependence), awareness of drug interactions and adverse drug reactions, and the proportion of patients who are aware of or unaware of the costs and benefits of the treatment.

Drug utilization research may generate hypotheses for further investigation by comparing drug utilization patterns and costs between different regions or time periods and by comparing observed patterns of drug use with current recommendations and guidelines for the treatment of a certain disease. These considerations should include both underuse and overuse of drugs. Determinants of use include user characteristics (e.g., sociodemographic parameters and attitudes towards drugs), prescriber characteristics (e.g., specialty, education, and factors influencing therapeutic decisions), and drug characteristics (e.g., therapeutic properties and cost).

A number of studies have addressed the factors that influence prescribing decision making, including education, advertising, colleagues, working circumstances, personality, control and regulatory measures, demands from society and patients, and cultural factors [41–43]. Some controversy exists concerning the relative impact of the various sources of influence on prescribing behavior, particularly the influence of pharmaceutical advertising. In studies of hospital practice the following factors have been stated to contribute to inappropriate prescribing: simple errors of omission; physician ignorance of cost issues in prescribing; failure to review medication orders frequently and critically; inability to keep up to date with developments in pharmacology and therapeutics; insulation of physicians and patients from cost considerations because of third-party coverage; and lack of communication between physicians and pharmacists [44].

Cultural factors are known to play a role in illness behavior and drug prescribing/consumption. A popular model describing cultural differences that may also influence drug prescribing is Hofstede's model of cultural dimensions. Five cultural dimensions are defined by which countries may be scored: power distance, individualism, masculinity, uncertainty avoidance, and long-term orientation. Power distance refers to the degree of hierarchy in a country and the extent to which the less powerful members of organizations and institutions accept and expect that power is distributed unequally. Individualism refers to the prevalence of the interests of an individual versus the group and the degree to which individuals are integrated into groups. Masculinity refers to a culture in which the emotional roles of the two genders are clearly separated. The assertive, competitive pole has been called "masculine" and the modest, caring pole "feminine." Uncertainty avoidance deals with a societal tolerance for uncertainty and ambiguity. It indicates to what extent a culture programs its members to feel either uncomfortable or comfortable in unstructured situations. Long-term orientation values are thrift and perseverance, while short-term orientation values are respect for tradition, fulfilling social obligations, protecting one's "face" [45]. It has been found that power distance and uncertainty avoidance are cultural dimensions associated with higher antibiotic use, suggesting that hierarchical societies use more antibiotics (difficulties dealing with authority), whereas more egalitarian societies use fewer, and in societies that tend to avoid uncertainty antibiotics have a defensive function (as the prescriber and the patient aim for certainty) [46,47].

Drug utilization research enables assessment of whether interventions to improve drug use had the desired impact, as well as the extent to which other factors influenced the pattern of use, including regulatory changes, reimbursement policy, pharmaceutical industry promotional activities, and others.

Intervention strategies aimed at improving prescribing behavior in hospital as well as primary care settings have been critically reviewed [48–51]. These are discussed in Chapter 19 and include dissemination of printed educational materials alone; multimedia warning campaigns; drug utilization audit followed by mailed or interactive feedback of aggregated results; group education through lectures or rounds; use of computerized reminder systems; use of opinion leaders to informally endorse or support specific behavior change interventions; one-to-one education initiated by a drug utilization expert; required consultation or justification prior to the use of specific drugs; and use of clinical guidelines.

Drug utilization research addresses the medical, social, and economic aspects of drug use [32,52]. Even when used appropriately, drugs have the potential to cause harm. However, a large proportion of their adverse effects is predictable and preventable [53,54]. Adverse drug reactions and drug nonadherence are important causes of hospital admissions in both adult and pediatric patients [54–56] (see also Chapter 38). Studies in the US have estimated that adverse drug events account for up to 28% of emergency department visits and 25% of ambulatory care encounters; up to 70% of these visits are deemed preventable [57]. Similar figures are also found in the UK and Sweden [54,58,59].

Many of these drug-related admissions may be preventable through the application of existing principles and data [60,61]. The situations that may lead to preventable adverse drug reactions and drug-induced illness include the use of a drug for the wrong indication; the use of a potentially toxic drug when one with less risk of toxicity would be just as effective; the concurrent administration of an excessive number of drugs, thereby increasing the possibility of drug–drug interactions (see Chapter 40); the use of excessive doses, especially for pediatric or geriatric patients; and continued use of a drug after evidence becomes

available concerning important toxic effects. Many contributory causes have been proposed: excessive prescribing by the physician; failure to define therapeutic endpoints for drug use; the increased availability of potent prescription and nonprescription drugs; increased public exposure to drugs used or produced industrially that enter the environment; the availability of illicit preparations; and prescribers' lack of knowledge of the pharmacology and pharmacokinetics of prescribed drugs [53]. Increased morbidity or mortality due to medication error [62] (see Chapter 41), poor patient adherence [63] (see Chapter 38), discontinuation of therapy [64–66], and problems in communication resulting from modern-day fragmentation of patient care are also to be considered (see Chapter 39).

Medication underdosing and underprescribing are often overlooked and can result in poor patient outcomes. The failure of physicians to prescribe an effective drug or effective doses for a treatable disease is a significant concern. For example, in a geographic area of Sweden with a higher suicide rate than average for the country, sales of antidepressant drugs were about half of those in other areas [67]. In the US, the underuse of beta-blockers in elderly patients with myocardial infarction was associated with an increased risk of death [20]. Other studies have documented significant underuse of antithrombotic drugs [21,68,69], lipid-lowering therapy [66,70–72], beta-blockers [22], aspirin [73], and thrombolytics [23] in patients with appropriate indications, but where outcomes were not assessed. In addition, underuse of beneficial medications may have other reasons, such as lack of access due to economic reasons or geographic access to the pharmacy and availability of prescription drugs [74–76].

Therapeutic practice, as recommended by relevant professional bodies, academic researchers, and opinion leaders, is initially based predominantly on data from premarketing clinical trials. However, the comparative effectiveness (i.e., the effectiveness of one medication compared to another medication in the real-world setting; see Chapter 26) and safety of new agents cannot be known with certainty until a drug has been on the market for many years or been extensively used. Complementary data from clinical experience and studies in the postmarketing period may result in changes in indication (e.g., a specific antibiotic no longer being a choice due to antimicrobial resistance), treatment duration (e.g., short-course antibiotic treatment of community-acquired pneumonia in children under 5 years of age), regimen (e.g., changes due to tolerance to oral hypoglycemic agents), precautions and contraindications (e.g., gastrointestinal bleeding with NSAIDs), and safety-based withdrawals [77,78]. For instance, when serious adverse reactions or special problems occur, particularly those that may lead to death or serious injury, a prominently displayed boxed warning, the so-called black box, is added to the US Food and Drug Administration (FDA) labeling of drugs or drug products. As therapy recommendations are updated through guidelines and other approaches, drug utilization studies must address the relationship between therapeutic practice as recommended and actual clinical practice [79].

Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

There are several methodologic issues in drug utilization research. Most of them are the same as for other pharmacoepidemiologic studies and are well described in other parts of the book. In this section, some specific issues of importance related to the different study designs in drug utilization are described, along with an overview of the available data sources.

Study Designs

There are many types of drug utilization studies. Research methods in drug utilization can be either *quantitative* or *qualitative*. Quantitative research deals with quantities; data are presented in numeric figures in categories or rank order and measured in various units. Quantitative research usually starts with a predefined hypothesis or theory, followed by data collection to provide an answer to the research questions formulated. Associations between variables and differences between different categories may be studied by using different statistical methods. Qualitative research, on the other hand, refers to the examination, analysis, and interpretation of observations for the purpose of discovering underlying meanings and patterns of relationships [80]. Qualitative studies include information that is not in numeric form collected through focus group discussions, open-ended questionnaires, in-depth interviews, and observations. Such studies may be used to explore the views of prescribers, dispensers, and patients in dealing with medicines. Consequently, they are important to gain a deeper understanding of various phenomena in drug utilization. Qualitative drug utilization studies are not further described in this chapter. For further reading on qualitative studies, there is a separate chapter on qualitative methods in drug utilization research in the handbook on drug utilization research [81].

The simplest quantitative drug utilization studies are descriptive. Such studies identify patterns or trends in drug utilization, without any attempt to draw conclusions about factors influencing drug use. Objectives of descriptive studies may be to quantify the present state, developmental trends, and time course of drug usage at various levels of the healthcare system, whether national, regional, local, or institutional. Routinely compiled drug statistics or drug utilization data that are the result of such studies can be used to estimate drug utilization in populations by age, sex, social class, morbidity, and other characteristics, and to identify areas of possible over- or underutilization. They also can be used as denominator data for calculating rates of *reported* adverse drug reactions in the context of spontaneous reporting systems (see Chapter 10); to monitor the utilization of specific therapeutic categories where particular problems can be anticipated (e.g., narcotic analgesics, hypnotics and sedatives, and other psychotropic drugs); or as markers for very crude estimates of disease prevalence (e.g., antiparkinsonian drugs for Parkinson's disease); to plan for drug importation, production, and distribution; and to estimate drug expenditures [34].

Descriptive drug utilization studies may also address quality of drug prescribing, dispensing, or use. In these studies, explicit predetermined criteria are created against which aspects of the quality, medical necessity, and appropriateness of drug prescribing may be compared. Drug use criteria may be based on such parameters as indications for use, daily dose, or length of therapy. Other possible criteria for poor drug prescribing include the failure to select a more effective or less hazardous drug if available, the use of a fixed combination drug when only one of its components is justified, or the use of a costly drug when a less costly equivalent drug is available [82]. In North America, these studies are known as *drug utilization review* (DUR) studies. For example, a large number of studies in North America have documented the extent of inappropriate prescribing of drugs, in particular antibiotics, and the associated adverse clinical, ecologic, and economic consequences [17–19].

Analytic drug utilization studies aim to gain a deeper understanding of the explanatory factors behind utilization patterns. The most robust analytic study designs are cohort studies or case–control studies. In traditional pharmacoepidemiology, such studies are used to assess the effectiveness or safety of drug therapy, where drug utilization is the exposure and clinical events constitute the outcome. In analytic drug utilization studies, drug exposure is the outcome

and the explanatory factors behind drug use constitute the exposure. The same methods for matching or confounder adjustment could thus be applied as for cohort and case–control studies used to assess safety or effectiveness. These methods are well described in other chapters of this book (see Chapter 43). Some examples of cohort studies in drug utilization include persistence studies, where discontinuation is the outcome to drug treatment [83–85]. Theoretically, case–control studies may also be conducted, selecting subjects on the basis of whether they have (or had) been prescribed or dispensed the drug of interest or not. An investigation of previous exposure to a factor might reveal whether there is an association between the drug utilization and previous exposure. Such studies are scarce in the literature, however.

In the absence of individual-level data to link drug utilization to other factors, ecologic studies could be conducted. In these studies, group-level data on dispensed or prescribed drugs are compared with other datasets, either for different geographic areas or population groups at a certain point in time or for the same population at different times. Some examples of ecologic studies in drug utilization research include the associations between antidepressant use and suicide [86]; respiratory medication prescribing, air pollution, and deprivation [87]; coronary heart mortality and statin use [88]; and unemployment rates and prescription drug utilization patterns [89]. Ecologic studies are simple to conduct, but they have limited value since no individual linkage has been conducted between exposure and outcome. Consequently, the correlations found in these studies cannot be interpreted as associations at the individual patient level.

Types of Data on Drug Utilization

A considerable amount of drug use data may be obtainable or is already available, the usefulness of which depends on the question at hand. All the data have certain limitations in their direct clinical relevance [90]. For quantitative studies, the ideal is a count of the number of patients in a defined population who ingest a drug of interest during a particular time frame, with a certain diagnosis or indication. The data available are only approximations of this for reasons that are described shortly, and thereby raise many questions about their presentation and interpretation.

Since most statistics on drug consumption were compiled for administrative or commercial reasons, the data were usually aggregated and expressed in terms of volumes or expenditure. First, data on drug utilization can be available as total costs or unit cost, such as cost per package, tablet, dose, or treatment course. Although such data may be useful for measuring and comparing the economic impact of drug use, these units do not provide information on drug exposure in the population. Moreover, data on expenditure are influenced by price fluctuations over time, distribution channels, inflation, exchange rate fluctuations, price control measures, and so on [91].

Volume data may be available from manufacturers, importers, or distributors as the overall weight of the drug that is sold or the unit volume sold; that is, the number of tablets, capsules, or doses sold. However, tablet sizes vary, making it difficult to translate weight into even the number of tablets. Prescription quantities also vary, so it is difficult to translate number of tablets into the number of exposed patients.

The number of prescriptions (either written or dispensed) has traditionally been one of the most frequently used measures in drug utilization studies. This measure may have some relevance in studies of medicines given for short treatment courses, such as antibiotics. For medicines used for chronic treatment the value is limited, since different patients receive a different number of prescriptions in any given time interval, and the amount allowed to prescribe or dispense on a prescription may vary substantially between countries. To translate

the number of prescriptions into the number of patients, one must divide by the average number of prescriptions per patient, or else distinctions must be made between first prescriptions and refill prescriptions. The former is better for studies of new drug therapy, but will omit individuals who are receiving chronic drug therapy. Additional problems may be posed by differences in the number of distinct drugs written in each prescription. Finally, it should be noted that these aggregate measures of prescribed or dispensed volumes represent approximate estimates of true consumption. The latter is ultimately modified further by the patients' actual drug intake; that is, their degree of adherence.

In the context of DUR, drug utilization data may be presented in the form of prescribing profiles for individual physicians or practices according to the number, monetary value, and even type of prescription ordered during a given time period. Pharmacies may also be ranked according to the number, cost, and type of prescription dispensed for similar intervals. However, these gross measures of prescription activity and drug use are limited in their capacity to reflect the wide spectrum of specific problems in prescribing. For example, they ignore problems such as the wrong drug for the indication, the wrong drug for the patient, the wrong dose, the wrong dosing interval, and the wrong duration of therapy. Also, one's deviation from the practices of the mean practitioner is not a good measure of one's "appropriateness" as a provider. Purely quantitative data characterizing prescribers as "high" or "low" may be driven, for example, by the number of patients seen by the physician and the type and severity of the patients' diseases. Data presented by pharmacy are even less informative, since patients may be dispensed prescriptions from an unknown range of different healthcare providers. However, for studies of medicine use in hospitals and studies of over-the-counter (OTC) products, pharmacy sales data may provide important information. Finally, it is important to emphasize that data on expenditures are not necessarily indicative of appropriateness, whether high or low relative to the mean.

In recent years, large patient-level prescription databases have become increasingly available [3]. They may contain all dispensed prescription drugs regardless of the reimbursement status and irrespective of who the prescriber is. These data are also closer to estimating actual drug exposure compared to prescribing data from electronic health records. However, there may be problems of poor sensitivity, with patients having other ways of receiving medications (e.g., drugs purchased abroad, OTC medicines, or drugs "borrowed" from relatives), or poor specificity, with patients not taking the drugs they have purchased. There are many useful methods available for studying drug utilization using individual data on dispensed prescriptions. Based on experiences from Denmark, Hallas and Støvring discussed three nonspecific analytic templates that could be applied to individual-level data on dispensed prescriptions [92]. Such methods include the ratio of prevalence odds to incidence rate to estimate the average duration for drug use, the Lorenz curve, and the waiting time distribution. The Lorenz curve expresses skewness in drug use. It shows the proportion of drug use that is accounted for by percentiles of drug users, ranked according to their volume of drug intake. It may express the extent of heavy users as well as sporadic small-volume users and may, for example, be used to screen for an unsuspected abuse potential of a drug. Figure 18.1 illustrates the use of insulin in a defined population: 50% of the users used 76% of the volume [92].

The waiting time distribution is a frequency distribution of first occurrences of drug use within a time window (Figure 18.2). It forms the basis for estimates of prevalence and incidence rate. Furthermore, it displays visual correlates of epidemiologic prescribing parameters such as

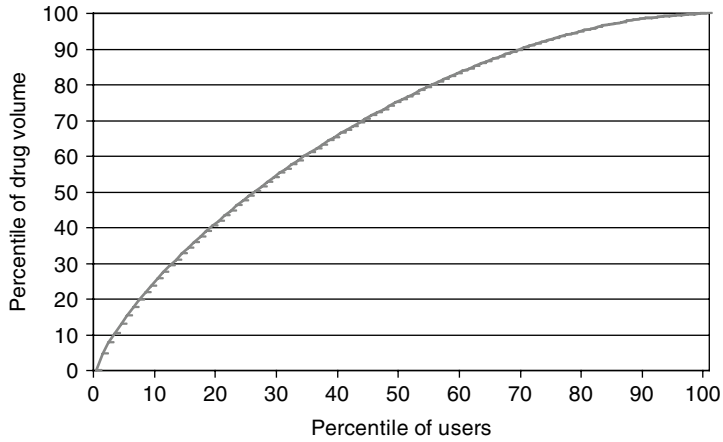


Figure 18.1 Lorenz curve for insulin use. The graph shows the proportion of insulin use that is accounted for by percentiles of insulin users, ranked according to their annual insulin consumption. Data from county of Funen, Denmark, 2003. *Source:* Hallas J, Støvring H. Templates for analysis of individual-level prescription data. *Basic Clin Pharmacol Toxicol* 2006; **98**(3): 260–5, Figure 1. Reproduced with permission of John Wiley & Sons.

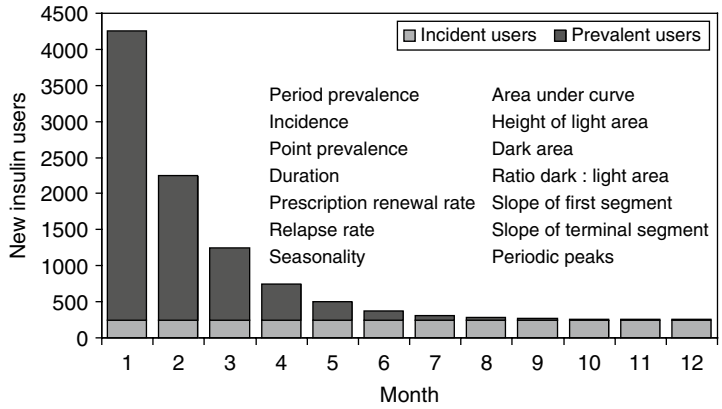


Figure 18.2 Visual correlates of various measures of insulin use. Hypothetical curve. Data from county of Funen, Denmark 2003. *Source:* Hallas J, Støvring H. Templates for analysis of individual-level prescription data. *Basic Clin Pharmacol Toxicol* 2006; **98**(3): 260–5, Figure 3. Reproduced with permission of John Wiley & Sons.

period prevalence, point prevalence, incidence rate, duration, prescription renewal rate, relapse of treatment, and seasonality.

From a quality-of-care perspective, to interpret drug utilization data appropriately, there is a need to relate the data to the reasons for the drug usage. Data on morbidity and mortality may be obtained from national registries (general or specialized); national samples where medical service reimbursement schemes operate; *ad hoc* surveys and special studies; hospital records; physician records; and patient or household surveys. "Appropriateness" of use must be assessed relative to indication for treatment, patient characteristics (age-related physiologic status, sex, habits), drug dosage (over- or underdosage), concomitant diseases (which might contraindicate or interfere with the chosen therapy), and the use of other drugs (interactions). However, no single source is generally available for obtaining all of this information. Moreover, because of incompleteness, the medical record may not be a very useful source of drug use data [93,94].

Generally agreed standards or criteria for appropriateness, based on currently available knowledge, are essential elements of the DUR process. These criteria must be based on scientifically established evidence; updated regularly according to new scientific evidence; explicitly

stated (to ensure consistency in evaluations); and applicable to a given setting. The development and standardization of these criteria are major undertakings. Finally, for DUR programs, even the strategy to be used to optimize one's intervention is still unclear.

Data Sources

A large variety of data sources can be used for drug utilization research. They may provide either primary or secondary data [95]. Primary data sources refer to original data collected by the investigator conducting the research for a specific purpose [96]. Secondary data sources include already collected data; that is, data which have not been usually generated for a specific research purpose but can be adapted to the analysis of a new research question [95].

Drug utilization studies have been conducted using a large variety of secondary data sources, including sales registries, procurement records, reimbursement/claims databases, medical records, pharmacy dispensing records, pharmacy stock records, disease-based registries, and population health surveys. The availability of such data varies substantially between countries, but there has been a large growth in access to them over time everywhere. The ongoing digitalization of healthcare brings further opportunities to access large amounts of clinical data for DUR. Some of these data may be unstructured and tricky to analyze, but new techniques and methods have been developed to address these challenges.

In the earlier editions of this book, we listed some diagnosis-linked and non-diagnosis-linked computer databases for drug utilization studies, as well as providing an overview of historical databases important for the development of pharmacoepidemiology and drug utilization research. In this edition, electronic databases are discussed in separate chapters (Chapters 11–14) and here we briefly give an overview of some types of sources that may be used. A more comprehensive overview of secondary data sources may be found in a recent textbook on drug utilization research [97].

Drug utilization data may be collected at any point in the pharmaceutical supply chain, starting with the manufacturer, passing through wholesalers and pharmacies, and ending with patients. Aggregate level sales data (from manufacturer/wholesaler/pharmacy) or purchases (from purchaser/payer, e.g., hospitals or community pharmacies) are regularly collected in most countries. Such data do not contain any information on number of patients, or any clinical data. Consequently, these data have limited value in analytic studies on the effectiveness and safety of medicine use. Still, they may be valuable in studies assessing quality of medicine use or to assess the effect on interventions in health systems. They may also be used in ecological studies comparing utilization patterns with other data, generating hypotheses to be studied with more robust study designs. At pharmacies, drug dispensing data may be recorded at an individual patient level. These data may subsequently be collected by insurers or reimbursement agencies. Such patient-level databases provide valuable sources for drug utilization, with and without linkage to other clinical information. Finally, healthcare providers are another important source of data on drug utilization studies, with information on drugs prescribed to patients available in their health records. Healthcare providers may also report information on selected drugs to disease-based patient registries.

Aggregated sales data have been used in drug utilization research for decades. Today, most countries keep some records of drug sales, at a regional or national level. These data can be obtained from health authorities [4,5] as well as from private companies such as IQVIA, a well-known commercial source of drug utilization data. An overview of European databases performed by Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) indicates

that aggregate sales data are widely collected across Europe [3,4]. In the US, the IQVIA National Sales Perspective database documents sales data for prescription drugs, OTC products, and some self-administered diagnostic products. Data collected include volume of dollars and quantities moving from manufacturers in various outlets within all states. In Canada, the IQVIA CompuScript database contains data on prescriptions sold from about two-thirds of Canadian retail pharmacies. Sales data collected by authorities or by companies such as IQVIA may be the only secondary source available to studies conducted in regions where other databases are not yet established or not accessible.

Dispensing data containing unique identifiers of patients are regularly collected at pharmacies. Since patients visit a pharmacy to fill their prescriptions, such data provide a more appropriate estimation on actual drug exposure than prescribing data from electronic health records. However, it is important to acknowledge that patients may also have other ways of obtaining medications (e.g., purchasing abroad, receiving in inpatient care, or "borrowing" from relatives), Furthermore, even though the patient claimed the prescription, it is uncertain whether the patient has taken the medicine as indicated by the prescriber. In Europe, the first patient-level research databases were established in the late 1960s to early 1970s [76]. These databases were based on prescriptions dispensed at the pharmacies and primarily used to study drug utilization. They were too small for drug safety studies. In the 1980s and 1990s larger databases were created, also based on administrative data, such as prescriptions dispensed at pharmacies, in Scotland [40], Denmark [98], Italy [99], and the Netherlands. An example of such a dispensing database is the Swedish Prescribed Drug Register, established in 2005, containing data with unique patient identifiers for the entire population of 10 million inhabitants for all dispensed prescriptions in ambulatory care [100]. This registry includes data on the patient (age, sex, personal identification number, place of residence), dispensed drug (Anatomic Therapeutic Chemical [ATC] classification code, defined daily dose [defined shortly] number, prescribed dose, package, reimbursement, date of prescribing and dispensing), prescriber (profession, specialty, workplace), and pharmacy (identifier, location). It can be linked to many other registers including outcome data and many quality registers with clinical data for different diseases. A recent review summarized the scientific output from the register after the first decade [101].

In many countries, third-party payers (public or private) collect medication data from pharmacies for billing purposes. Drug reimbursement/claims databases typically contain unique patient identifier, prescriber (either individual prescriber and/or clinic/practice), pharmacy dispensing the drug, drug name and ATC code, strength, dosage form, quantity dispensed, date of prescription and dispensation, days' supply, as well as patient co-payment and total drug expenditure. In some countries reimbursement data also include information on diagnosis, as an International Classification of Diseases (ICD) code is written on prescriptions. Reimbursement databases are commonly used in DUR. A limitation, though, is that they may only contain those drugs that are reimbursed. Furthermore, they are sensitive to changes in co-payment over time. There are also differences between health systems in regulations and reimbursement and, consequently, the amount of information documented in these databases varies. In some countries, prescription drugs are funded only for selected groups of the population (e.g., the elderly) [102,103], and various other public and private prescription plans may be used for other population groups.

In countries with many different insurers (e.g., in the US), it may be difficult to follow people over time, since patients move in and out as their insurance eligibility changes [104]. In the

US, Medicaid medical and pharmaceutical billing data have been available for drug utilization studies for many years. With the disadvantaged and disabled population included in Medicaid, however, the generalizability of the results is a potential concern, especially for such descriptive studies. In contrast, in many European and some Asian countries with national health systems, insurance coverage can be close to 100% of the population, thus providing a complete picture of medicine use in these countries [4,105].

Large databases are also derived from electronic health records. The key advantage of these databases is that they contain clinical data such as diagnoses. One example is the General Practice Research Database® (GPRD®) in the UK (see Chapter 14), which is based on medical records from general practitioners (GPs). Hundreds of GPs contribute anonymized patient information to a central database, that now contains millions of patients. Included are prescriptions issued by the GP but with no information from the pharmacy (compliance/adherence). All these databases were primarily used for drug safety studies, but have also been used to study drug utilization [96,106].

Another example is the Integrated Primary Care Information (IPCI) database, established at Erasmus University in the Netherlands. It consists of the computer-based patient records of 150 general practitioners. To date the database has accumulated data on approximately 500 000 patients. The records are coded to ensure the anonymity of patients; data include patient demographics, symptoms (in free text), diagnoses (based on the International Classification for Primary Care and free text), clinical examination findings, referrals, laboratory test results, hospitalizations, and physician-linked drug prescriptions and dosage regimen (but no information from the pharmacy on compliance/adherence).

The National Disease and Therapeutic Index (NDTI), by IQVIA, is an ongoing study of physician prescribing that is conducted mainly for use by pharmaceutical companies for marketing [107]. This study employs a rotating sample of office-based physicians, who record all patient encounters and corresponding "drug mentions" for two-day periods four times a year. A special prescription form is used to collect information on the drug (specific product, dosage form, new vs. continuing therapy), patient characteristics (sex), prescriber (specialty, location, region), type of consultation (first versus subsequent), concomitant drugs and diagnoses, and the desired pharmacologic action [15]. Data have been made available to academic researchers (for a fee) and the FDA [15]. Although useful for studies of prescribing, longitudinal patient-specific studies are not possible with this database.

Currently Available Solutions

DUR studies are activities aimed at the detection and quantification of problems. They should be distinguished from DUR *programs*. DUR studies are usually one-time projects, not routinely conducted. They provide only minimal feedback to the involved prescribers and, most importantly, do not include any follow-up measures to ascertain whether any changes in drug therapy have occurred. A DUR program, on the other hand, is an intervention in the form of an authorized, structured, and *ongoing system* for improving the quality of drug use within a given healthcare institution. The quality of drug prescribing is evaluated by employing predetermined standards for initiating administrative or educational interventions to modify patterns of drug use that are not consistent with these standards. The measurement of the effectiveness of these interventions is ideally an integral part of the program [108,109].

In the US, DUR programs (commonly known in hospitals as DUE programs) are part of the quality assurance activities required by

Medicaid–Medicare regulations, the Joint Commission, the former Professional Standards Review Organizations (PSRO), and Section 4401 of the Omnibus Budget Reconciliation Act of 1990 [109]. In Europe, DUR programs have been proposed as periodic "therapeutic audits" performed at various levels (patient, prescriber, hospital, county, municipality, country, and groups of countries), assessing not only the clinical consequences of drug utilization, but also the social and economic consequences. These studies are to be followed by whatever feedback is felt to be necessary and appropriate to effect changes in therapeutic practices [6,110,111]. Most commonly, these therapeutic audits have been based on aggregate data analysis of medicines consumption at a national level and interventions, usually regulatory or informational and educational, and are aimed accordingly at whole populations or subgroups, rather than specific individuals. Despite their widespread implementation in the US, the effectiveness of DUR programs in reducing prescribing errors and improving patient outcomes remains to be established (as discussed later).

Units of Measurement

The defined daily dose (DDD) method was developed in response to the need to convert and standardize readily available volume data from sales statistics or pharmacy inventory data (quantity of packages, tablets, or other dosage forms) into medically meaningful units, to make crude estimates of the number of persons exposed to a particular medicine or class of medicines [112].

The DDD method is useful for working with readily available gross drug statistics; allows comparisons between drugs in the same therapeutic class and between different healthcare settings or geographic areas, and evaluations of trends over time; and is relatively easy and inexpensive to use. The method is firmly established in Europe and is increasingly used by researchers in other regions [113–118]. A WHO manual on drug utilization research provides an overview of the method [119]. Guidelines for classifying medicines and their assigned DDDs are updated annually by the WHO Collaborating Centre for Drug Statistics Methodology (www.whocc.no).

The DDD is a technical statistical unit defined as the assumed average daily maintenance dose for a drug for its main indication in adults. It is only a measurement unit and does not necessarily reflect the prescribed or recommended dose. To enable comparison of drug use in DDDs the information has to be presented with an adequate denominator; that is, the population that was exposed to a drug. Ambulatory drug use is commonly expressed as DDDs per 1000 inhabitants per day. For chronically used drugs, it can be interpreted as the proportion of the population that receives treatment with a particular medicine on any given day. For example, if the use of a drug is measured as 30 DDDs/1000 inhabitants/day in a given country, this indicates that around 3% of the country's population receives that drug daily. Sometimes better estimates can be given by adjusting the denominator for a target population (e.g., for oral contraceptives the denominator is females below 45 years of age). For use in hospital settings, the unit is expressed as DDDs per 100 bed-days (adjusted for occupancy rate); it suggests the proportion or percentage of inpatients that receive a DDD in a day. For example, 30 DDDs/100 bed-days indicates that 30% of patients in a day receive a certain drug. For medicines that are used in the outpatient setting for short-term periods, such as antimicrobials, the unit is expressed as DDDs per inhabitant per year; this provides an estimate of the number of days for which each person is treated with a particular medication in a year. For example 8 DDDs/inhabitant/year indicates that every inhabitant is on average treated with that drug for 8 days in a year.

The DDD method has been useful in describing and comparing patterns of drug utilization [1,2,111], providing denominator data to estimate reported adverse drug reaction rates [120], performing epidemiologic screening for problems in drug utilization [111], and monitoring the effects of informational and regulatory activities [113,121]. It has also been used to study variations in antimicrobial utilization [9,122,123], as well as antimicrobial utilization and its correlation with antimicrobial resistance in outpatient [123,124] and inpatient settings in Europe [125], and to report on sustained reduction of antibiotic use and low bacterial resistance with implementation of a multidisciplinary, coordinated national antimicrobial and rational use program [126].

The European Surveillance of Antimicrobial Consumption Network, ESAC-Net (formerly ESAC), collects and analyzes data on antimicrobial consumption from European Union (EU) and European Economic Area/European Free Trade Association (EEA/EFTA) countries, using the DDD methodology. The data on community and hospital antimicrobial consumption are publicly available from ESAC-Net [127].

The DDD method should be used and interpreted with caution. The DDD is not a recommended or a prescribed dose, but a technical unit of comparison; it is usually the result of literature review and available information on use in various countries. Thus, the DDDs may be high or low relative to actual prescribed doses. Moreover, the DDDs refer to use in adults. Since children's doses are substantially lower than the established DDDs, if unadjusted this situation will lead to an underestimation of population exposures, which may be significant in countries with a large pediatric population. Although pediatric DDDs have also been proposed [128], the concept and its applicability have not been incorporated into the WHO method [119]. Finally, DDDs do not take into account variations in adherence.

The prescribed daily dose (PDD) is another unit, developed as a means to validate the DDD. The PDD is the average daily dose prescribed, as obtained from a representative sample of prescriptions [129]. Problems may arise in calculating the PDD because of a lack of clear and exact dosage indication in the prescription, as is often the case with the prescribing of insulin. Prescriptions for chronic therapy, as in the case of insulin, may be refilled many times and the dosage may be altered verbally between prescribing events [130]. For certain groups of drugs, such as oral antidiabetics, the mean PDD may be lower than the corresponding DDD. Up to twofold variations in the mean PDD have been documented in international comparisons [129]. Higher PDDs have been observed in the US relative to Sweden for commonly prescribed drugs, such as hydrochlorothiazide, diazepam, and oxazepam [131]. In studies assessing whether antidepressants increase the risk of suicide, a refined person-year of use estimate was obtained from adjusting the DDD by the average PDD for individual antidepressants [132]. Although the DDD and the PDD may be used to estimate population drug exposure "therapeutic intensity," the method is not useful to estimate incidence and prevalence of drug use, or to quantify or identify patients who receive doses lower or higher than those considered effective and safe.

The Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America (IDSA/SHEA) have recommended days of therapy (DOT) for expressing antimicrobial drug use [133]. DOT is the number of days when at least one dose of a medication was administered irrespective of dose or route of administration. They are not impacted by dose adjustments and can be used in both adult and pediatric populations. Similar to PDDs, expressing drug use in the number of DOTs requires patient-level use data, which may not be feasible at every facility.

Drug and Disease Classification Systems

The Anatomic Therapeutic Chemical classification system is generally used in conjunction with the DDD method [112,119]. It was originally

developed by the Norwegian Medicinal Depot, which became a WHO Collaborating Centre for Drug Statistics Methodology; the center is now located at the Norwegian Institute of Public Health (www.whocc.no). The ATC system is based on the main principles of the Anatomical Classification system developed by the European Pharmaceutical Market Research Association (EPhMRA) and the International Pharmaceutical Market Research Group (IPMRG).

The ATC system consists of five hierarchical levels: a main anatomical group, two therapeutic subgroups, a chemical-therapeutic subgroup, and a chemical substance subgroup. The coding of furosemide preparations is used to illustrate the ATC classification structure in Table 18.1. The first three levels are modifications of the three-level EPhMRA and IPMRG classification system. The fourth and fifth levels are extensions that are developed and updated by the WHO Collaborating Centre for Drug Statistics Methodology. Ongoing discussions aim to identify differences in the two classification systems and harmonize the first three levels. Statistics reported with the ATC system should not be directly compared with figures prepared with the EPhMRA system.

Medicinal products are classified according to the main therapeutic indication for the principal active ingredient. Most products are assigned only one ATC code. However, some active medicinal substances may have more than one ATC code, if the drug has different uses at different strengths (acetylsalicylic acid as a platelet aggregation inhibitor and as an analgesic–antipyretic), dosage forms (timolol to treat hypertension and to treat glaucoma), or both (medroxyprogesterone for cancer therapy and as a sex hormone). Prednisolone is an example of a drug that has six different codes. Fixed-dose combination products pose classification difficulties. For example, a combination product that contains an analgesic and a tranquilizer is classified as an analgesic, even though it also contains a psychotropic substance. Because the ATC codes and DDDs may change over time with regular revisions, researchers must carefully document which version of the classification and DDD assignment is used, so that the resulting drug statistics may be adequately interpreted [134].

The European Drug Utilization Research Group (EuroDURG), formerly the WHO Drug Utilization Research Group and currently an association of European national Drug Utilization Research Groups, and the International Society of Pharmacoepidemiology Special Interest Group in Drug Utilization Research (ISPE SIG DUR) recommend the use of the ATC classification system for reporting drug consumption statistics and conducting comparative drug utilization research [130]. The WHO International Drug Monitoring Program uses the system for drug coding in adverse drug reaction monitoring (www.who-umc.org). Some developing countries also use the ATC system to classify their essential drugs [135,136]; this may eventually lead to the preparation of annual drug utilization statistics [137].

In the US, the Iowa Drug Information System (IDIS) is a hierarchical drug coding system that is based on the three therapeutic categories of the American Hospital Formulary Society (AHFS), to which a fourth level was added to code individual drug ingredients [138]. The IDIS code has eight numeric digits, two digits per level (see Table 18.1). This coding system was used in the Established Populations for Epidemiologic Studies of the Elderly survey [138]. Other coding systems such as the National Drug Code and the Veterans' Administration Classification [139] do not provide unique codes for drug ingredients.

In the UK, British National Formulary (BNF) codes are widely used for drug utilization studies. The BNF provides monographs for drugs available in the UK. The numbering system is produced by NHS Prescription Services, part of the NHS Business Services Authority in England [140].

Table 18.1 Anatomic Therapeutic Chemical (ATC) and Iowa Drug Information System (IDIS) classification and coding structures for furosemide.

ATC Classification (C03CA01)					
C	Cardiovascular System (First level, main anatomical group)				
	03	Diuretics (Second level, main therapeutic group)			
		C	High-ceiling diuretics (Third level, therapeutic subgroup)		
			A	Sulfonamides, plain (Fourth level, chemical therapeutic subgroup)	
				01	Furosemide (Fifth level, chemical substance)
IDIS Classification (40280401)					
40	Electrolyte Solutions (First level, main therapeutic group)				
	28	Diuretics (Second level, therapeutic subcategory)			
		04	Loop-diuretics (Third level, therapeutic subcategory)		
			01	Furosemide (Fourth level, chemical substance)	

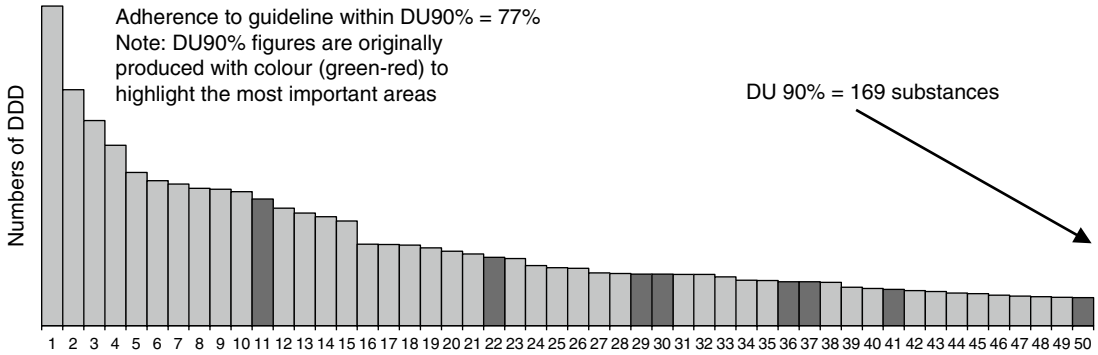
The International Classification of Diseases is a system of diagnostic codes for classifying diseases and other health problems. The ICD is published by the WHO and used worldwide in morbidity and mortality statistics, drug reimbursement systems, and automated decision support in healthcare. The system includes categories relating to medicinal substances, but in the context of adverse outcomes, and often in quite broad terms. It does not include codes suitable for recording and classifying drug utilization [141].

The Systematized Nomenclature of Medicine (SNOMED) provides a core general terminology for use in various medical fields. SNOMED clinical terms (CT) contain more than 311 000 active concepts in clinical settings, organized in different hierarchies. An individual number represents each concept, and several concepts can be used in combination to describe a complex condition. Clinical finding/disorder and procedure/intervention are examples of the main levels in SNOMED CT. Substance and pharmaceutical/biologic product are also in the main levels. The pharmaceutical/biologic product hierarchy was introduced as a top-level hierarchy in order to distinguish drug products from their chemical constituents (substances). It contains multiple levels of granularity, used to support a variety of purposes, including electronic prescribing and formulary management (www.ihtsdo.org).

Quality Indicators

Drug utilization studies assessing the quality of drug prescribing involve the use of various types of quality indicators [142]. These may be defined as "A measurable element of prescribing performance for which there is evidence or consensus that it can be used to assess quality, and hence in changing the quality of care provided." Quality indicators for drug utilization may be classified based on the amount of clinical information incorporated in the indicator. *Drug-oriented indicators* focus solely on the drugs prescribed, dispensed, or consumed. The simplest approaches only require aggregate data on the volume and expenditure of prescribed or dispensed drugs. Such data could be presented as time trends or top-ten lists and used as a catalyst to stimulate discussion around areas for improvement in drug therapy. Simple drug-oriented indicators can be constructed to compare practices, clinics, or regions. Such drug-oriented indicators are based on drug data alone and can be used irrespective of the indication for which the drug is prescribed. The most commonly used drug-oriented indicator is the ratio between different drugs. Some examples include the ratio of COX-2 inhibitors to all NSAIDs measured in DDDs, the ratio of simvastatin to statins, and the ratio of angiotensin-converting enzyme (ACE) inhibitors to all renin-angiotensin drugs [143–146]. Other types of drug-oriented indicators have been focused on inappropriate drugs in children and the elderly [147–150].

Another approach analyzed the number of drugs that accounted for 90% of drug utilization (DU90%) and the percentage of these drugs that were included in evidence-based guidelines [151]. The first studies on the DU90% method used the guidelines on rational drug use issued by the regional Drug & Therapeutics Committee in Stockholm, Sweden [152]. (See Figures 18.3 and 18.4.) These guidelines consist of approximately 200 medicines recommended as first-line choices for the treatment of common diseases. The 90% level was arbitrarily selected to focus on the bulk of prescribing, yet allow some degree of individual variation. The number of different products in the DU90% segment varied between 117 and 194 among 38 primary healthcare centers in Stockholm; adherence to the guideline varied at between 56% and 74%. The Swedish Medical Quality Council has recommended the DU90% method for assessing quality in drug prescribing. Using this method, researchers in the Netherlands did not find any



	SUBSTANCE	(DDD)	DDD	% TOT	Rx	COST	COST/DDD
1	**Acetylsalicylic acid**	1 tablet	39,894,782	4.9%	650,808	22,995,814	0.58
2	**Simvastatin**	30 mg	29,455,125	3.6%	438,802	26,731,380	0.91
3	**Enalapril**	10 mg	25,632,413	3.2%	296,329	18,111,103	0.71
4	**Furosemide**	40 mg	22,513,352	2.8%	409,630	18,091,910	0.80
5	**Omeprazol**	20 mg	19,140,338	2.4%	399,298	30,122,076	1.57
6	**Cyanocobalamin**	1 mg	18,125,259	2.2%	319,737	11,711,704	0.65
7	**Amlodipine**	5 mg	17,685,421	2.2%	165,634	10,209,168	0.58
8	**Metoprolol**	0.15 g	17,160,653	2.1%	498,845	72,421,602	4.22
9	**Levothyroxine sodium**	0.15 mg	17,030,980	2.1%	405,353	23,094,330	1.36
10	**Ramipril**	2.5 mg	16,743,688	2.1%	95,412	8,323,952	0.50
11	Felodipine	5 mg	15,807,331	2.0%	177,725	13,901,701	0.88
12	**Candesartan**	8 mg	14,695,169	1.8%	118,979	59,427,671	4.04
13	**Zopiclone**	7.5 mg	14,059,603	1.7%	426,875	16,073,585	1.14
14	**Paracetamol**	3 g	13,597,335	1.7%	621,717	35,916,832	2.64
15	**Citalopram**	20 mg	13,065,325	1.6%	266,993	12,672,896	0.97
16	**Sertraline**	50 mg	10,178,236	1.3%	119,262	11,241,414	1.10
17	**Hydroc.thiazide + amiloride**	*	10,145,923	1.3%	120,927	5,066,822	0.50
18	**Calcium combinations**		10,047,562	1.2%	206,568	24,983,957	2.49
19	**Propiomazine**	25 mg	9,729,948	1.2%	174,248	12,173,527	1.25
20	**Metformin**	2 g	9,297,143	1.2%	152,190	18,415,868	1.98
...							
169							
DU 90%	**1 - 169**		727,456,526	90.0%	14,034,056	2,648,199,198	3.64
170	- 853		80,552,999	10.0%	2,699,117	2,370,147,348	29.42
TOTAL	**1 - 853**		**808,009,524**	**100.0%**	**16,733,173**	**5,018,346,546**	**6.21**

Bold = in guideline
**** = Different DDDs for various routes of administration***
Medicines without DDD excluded (455, corresponding to 511 million SEK)

Figure 18.3 DU90% (number of substances accounting for 90% of the volume in DDDs) in Stockholm Healthcare Region in 2009. Dark gray, nonrecommended drugs; DDD, defined daily dose; DU, drug utilization. *Source:* Gustafsson LL, Wettermark B, Godman B, *et al*. The "Wise List": a comprehensive concept to select, communicate and achieve adherence to recommendations of essential drugs in ambulatory care in Stockholm. *Basic Clin Pharmacol Toxicol* 2011; **108**(4): 224–33, Figure 4. Reproduced with permission of John Wiley & Sons.

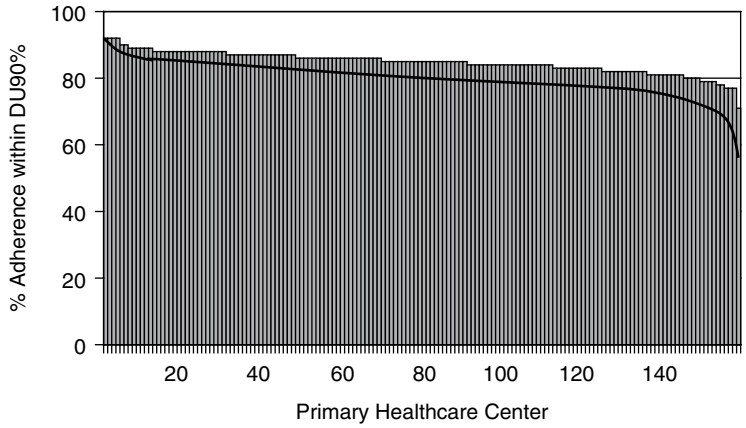


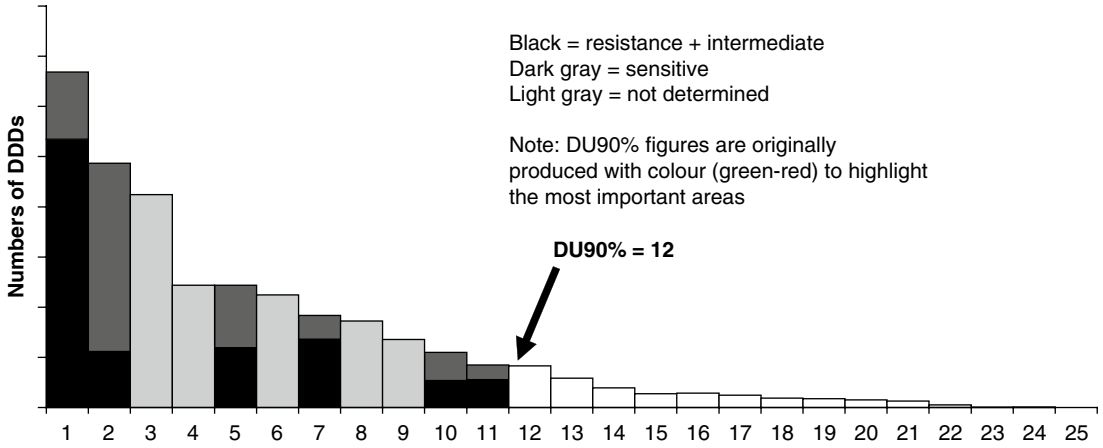
Figure 18.4 Adherence to "Wise List" recommendations for 156 primary healthcare centers for prescriptions dispensed in 2009. The thick black line equals the adherence range for the same practices in 2003. Observe that the order of the practices may differ between the two years. *Source:* Gustafsson LL, Wettermark B, Godman B, *et al*. The "Wise List": a comprehensive concept to select, communicate and achieve adherence to recommendations of essential drugs in ambulatory care in Stockholm. *Basic Clin Pharmacol Toxicol* 2011; **108**(4): 224–33, Figure 5. Reproduced with permission of John Wiley & Sons.

association between different levels of performance in pharmacotherapy audit meetings and quality of prescribing for seven drug classes. They suggested that for certain drug classes, such as antidepressants, duration of treatment may be more relevant for quality prescribing than using the drug of first choice in the guidelines; for diabetes, co-medication with statins may be more important than the number of different oral antidiabetics; and for obstructive airway diseases, concomitant use of corticosteroids may be more a more appropriate criterion than choices within the guidelines [153].

DU90% has also been used, for example, to compare NSAIDs prescribing in Denmark, Italy, Croatia, and Sweden [7,8], antibiotics in Denmark and Italy 199], and general intensive care unit antibiotic prescribing and cost patterns in Israel [154], and to assess the effect of financial incentives linked to self-assessment of prescribing patterns in Swedish primary care [155]. In Stockholm, the DU90% profile was useful in following up adherence to recommendations of essential drugs in ambulatory care for 15 years [152,156]. Furthermore, integrating resistance to the antibiotic DU90% profiles showed striking figures on the use of antibiotics that were cheap but resistant rather than effective and more expensive [157]. (See Figure 18.5.)

Access to patient-level data enables the construction of more clinically relevant drug-oriented indicators linking different drugs to each other or over time. Some examples of approaches include the identification of inappropriate or interacting drugs (questionable combinations) prescribed to individual patients [60,61,150,158,159] or co-prescribing of beta-adrenoreceptor antagonists and agonists [160]. They may also be used to identify in which order drugs are initiated to patients, for instance the proportion of patients initiated on angiotensin receptor blockers previously dispensed ACE-inhibitors [161].

Disease-oriented indicators include information on drugs linked to the diagnosis or healthcare problem. They may either indicate to what extent patients are being treated with the recommended drugs for a certain condition, or to what extent drugs are avoided for patients with conditions for which the drugs should not be used.



	SUBSTANCE	(DDD)	DDD	% TOT	DDD/100 BED-DAYS	COST/DDD (rubles)	% RESIST
1	Gentamicin	0.24 g	6 687	20%	3,1	5	80%
2	Cefazolin	3 g	4 865	15%	2,3	41	23%
3	Amoxicillin + clav. acid	1 g	4 242	13%	2,0	105	nd
4	Ampicillin + Oxacillin	8UD (2g)	2 438	7%	1,1	13	nd
5	Benzylpenicillin	3.6 g	2 435	7%	1,1	129	49%
6	Pefloxacin	0.8 g	2 245	7%	1,1	6	nd
7	Ampicillin	2 g	1 838	6%	0,9	10	74%
8	Metronidazole	1.5 g	1 726	5%	0,8	50	nd
9	Lincomycin	1.8 g	1 357	4%	0,6	5	nd
10	Amikacin	1 g	1 099	3%	0,5	143	49%
11	Ciprofloxacin	1 g	845	3%	0,4	400	66%
12	Pipemidic acid	0.8 g	828	2%	0,4	22	nd
DU 90%	**1 - 12**		30 603	92%	14,4		
13 - 25			2 477	8%	1,2		
TOTAL	**1 - 25**		**33 080**	**100%**	**15,5**		

Figure 18.5 DU90% profile for "antibacterial for systemic use" (J01, ATC/DDD classification) in university hospital N2 of St Petersburg, Russia in 2003. The 12 antibiotics are ranked in order of number of defined daily doses (DDDs) corresponding to 90% of use (data from the hospital pharmacy). The black parts correspond to the percentage of resistance (resistance + intermediate) for the antibiotics, and the dark gray to the percentage of sensitivity. Antibiotics not tested for bacterial sensitivity are light gray. nd, not determined. *Source:* Goryachkina K, Babak S, Burbello A, Wettermark B, Bergman U. Quality use of medicines: a new method of combining antibiotic consumption and sensitivity data – application in a Russian hospital. *Pharmacoepidemiol Drug Saf* 2008; **17**(6): 636–44.

Patient-oriented indicators include information on drugs linked to individual clinical characteristics of the patient, such as the severity of the disease and whether a certain treatment is suitable for a specific patient [160].

Although the use of health insurance databases has been reported in countries outside North America, Europe, and Asia [31,162–164], medical and pharmaceutical databases are generally not available in most low and middle income countries. An approach based on the use of standardized criteria (indicators) to measure changes in medicines prescribing, dispensing, and patient care was developed in the

Table 18.2 World Health Organization/International Network for the Rational Use of Drugs drug use indicators.

Core indicators
Prescribing indicators
Average number of drugs per encounter
Percentage of drugs prescribed by generic name
Percentage of encounters with an antibiotic prescribed
Percentage of encounters with an injection prescribed
Percentage of drugs prescribed from essential drugs list or formulary
Patient care indicators
Average consultation time
Average dispensing time
Percentage of drugs actually dispensed
Percentage of drugs adequately labelled
Patient's knowledge of correct dosage
Facility indicators
Availability of copy of essential drugs list or formulary
Availability of key drugs
Complementary indicators
Percentage of patients treated without drugs
Average drug cost per encounter
Percentage of drug costs spent on antibiotics
Percentage of drug costs spent on injections
Prescription in accordance with treatment guidelines
Percentage of patients satisfied with care they received
Percentage of health facilities with access to impartial drug information

Source: How to Investigate Drug Use in Health Facilities: Selected Drug Use Indicators. EDM Research Series No. 007. Reproduced with permission of WHO.

early 1990s by INRUD and the WHO [165]. The approach has facilitated the study of drug utilization in developing countries. It includes recommendations on minimum sample sizes, sampling methods, and data collection techniques, depending on study objectives. The method recommends 12 core indicators and 7 complementary indicators to study drug use in health facilities (Table 18.2). These indicators can be used to describe prescribing practice [166], for conduct monitoring and supervision [167], and to assess the impact of interventions [168–170]. The WHO has compiled indicator results and other findings reported in studies conducted in 97 developing and transitional countries between 1990 and 2006 [171].

INRUD has also developed simple low-cost indicators to measure adherence to antiretroviral (ARV) treatment in resource-poor settings. Adherence measures derived from dispensing data in pharmacy records, self-report data in medical records, and attendance logs predicted key clinical outcome related to individual patient treatment success, and were feasible to collect [172]. The four indicators were percentage of patients with self-reported full adherence, percentage of days covered by ARVs dispensed, percentage of records with a 30-day gap in ARVs dispensed, and percentage of patients who attended within 3 days of the scheduled appointment. These indicators allow assessment and comparison of programs and facilities, and monitoring and evaluation of interventions.

Intervention Strategies Based on Drug Utilization Data

Numerous studies have described interventions aimed at improving prescribing by the use of drug utilization data obtained from drug utilization studies, and are discussed further in Chapter 19. Two intervention strategies may illustrate different approaches to the use of drug utilization data available from computer databases of office practice.

In a randomized clinical trial, Avorn and Soumerai [173] used Medicaid data to identify physicians who were prescribing drugs that were assessed as inappropriate (based on considerations of documented efficacy, relative efficacy, and relative cost). These physicians were targeted for educational or information activities, as either face-to-face contacts or written drug information. Schaffner *et al.* [174] and Ray *et al.* [175] used a similar approach in another controlled intervention study, comparing different strategies aimed at modifying physician prescribing behavior: written drug information versus personal visits by pharmacists versus personal visits by physician educators. These two studies demonstrated the efficacy of face-to-face methods in improving drug prescribing.

The second approach uses claims data to perform computerized screening for patients who may be at increased risk for drug-induced illness, using patient-specific medical and drug histories [176–178]. Health professionals then evaluate profiles of patients with possibly inappropriate drug use. If drug use is indeed considered inappropriate, a letter is sent to the prescriber providing a profile of the patient's relevant computerized claims record and a warning of the potential for drug-induced disease. Often the problem is a concomitant drug or diagnosis of which the prescriber was unaware. This approach is obviously much less expensive than the face-to-face approach. Using before-and-after comparisons, a significant reduction in drug-induced hospitalizations has been noted [177]. However, the interpretation of these results is hampered by the use of a nonexperimental design. Other authors have found no effect on measures of prescribing or on patient outcomes [179]. A simultaneously controlled trial is needed to adequately assess the value of this approach.

Many other studies have described intervention strategies based on providing drug utilization data feedback, alone or in combination with printed material and/or other "educational strategies," for example group discussions, lectures, seminars, or personal visits by "experts." The results from these studies are conflicting. Some suggest that methods that involve only feedback of drug utilization data or audit results are ineffective. Others suggest a transient effectiveness for those that combine the use of drug utilization review data with group discussions, lectures, and visits by experts. However, these are difficult to interpret because of limitations in their research designs [180].

Conceptually, DUR programs are aimed at the improvement of medical care and cost containment. However, in practice traditional approaches have focused on control of the abuse or overuse of drugs, polypharmacy, or patients obtaining prescriptions from many different prescribers. Moreover, most DUR studies have emphasized process measures of quality of care, for example the use of clinical laboratory tests to monitor for adverse effects during chloramphenicol or aminoglycoside therapy. The approach described by Strom *et al.* [176], Lee Morse *et al.* [177], and Groves [178] was a significant advance in DUR programs, as it was primarily aimed at improving measurable patient outcomes. Also, it does not impose arbitrary restrictions on drug use, potentially impairing patient care, but seeks to reduce costs by improving patient care. In seeking to reduce the financial impact of drug use, it does not focus on the drug costs themselves, but on the effects of the drugs. By reducing the need for medical care through the beneficial effects of drugs, or by increasing the need for remedial medical care because of drug toxicity, pharmaceuticals can have a financial impact on the healthcare system that is much larger than the cost of the drugs themselves. (This is discussed more in Chapter 35.)

Despite their appeal, the effectiveness of DUR programs remains to be established. A study of six Medicaid programs failed to identify an effect of retrospective DUR on the rate of potential prescribing errors and rate of all-cause or specific-cause hospitalizations [179]. Another study did not find effects of two state prospective DUR interventions on the frequency of drug problems, utilization of prescription drugs and other health services, and clinical outcomes [181].

The Future

Opportunities

Drug utilization research is rapidly expanding in all countries across the globe, from the early descriptive studies to advanced studies combining different data sources to further understand medicine use in the population. In the early days, drug utilization studies were suggested to focus on the medical, social, and economic

aspects of drug utilization [2]. Medical aspects included potential inappropriate prescribing in different groups such as children and the elderly. Social aspects included attitudes to medicines and health, drug abuse and dependence, and their causes and trends, as well as socioeconomic inequities. Economic aspects included drug prices and expenditure for generics and brands, as well as allocation of resources (money, personnel, facilities) to drugs and other aspects of healthcare. All these aims are still relevant for future drug utilization research. However, the types of drugs in focus will differ, with 42% of the substances in drug development being biologics, compared to 8% on the market today [182]. The growing pressures on all healthcare systems with aging populations, rising patient expectations, stricter clinical targets, and expensive new medicines will further increase the need for drug utilization studies to monitor that resources are used wisely and that new medicines are prescribed to those who may benefit most from them.

From a public health perspective, the observed differences in national and international patterns of drug utilization require much further study. The medical consequences as well as the explanations for such differences are still not well documented. The increasing availability of patient-level databases on dispensed medicines will facilitate studies of the incidence and prevalence of medicine use, as well as more sophisticated studies on patterns of use, including drug combinations, dosing regimens, and persistence to drug therapy. These databases contain or may be linked to diagnoses and other clinical data that facilitate drug utilization studies, where drug utilization can be understood in its clinical context. The ongoing digitalization of healthcare will further increase access to large amounts of data for drug utilization research. Considerable amounts of healthcare data are generated every day, some of it from data sources to a large extent unexplored or unused in drug utilization research, such as clinical records systems, mail traffic, social media, and various devices.

Numerous studies have addressed the factors influencing drug prescribing. However, the relative importance of the many determinants of appropriate prescribing is still to be adequately elucidated. Further research is needed to better define to what degree and which determinants of inappropriate prescribing are susceptible to modification, and what might be an appropriate mix of interventions to achieve optimal impact. Although regulation is effective, it is not possible to regulate all aspects of the clinical decision-making process to ensure optimal drug prescribing [183]. Other approaches in addition to educational and informational measures are being explored. It is also important to emphasize the growing role of the patient in drug utilization research, both as a source of information to understand how drugs are used in reality, and also as a partner in designing and conducting research.

There is a need too for many more intervention studies targeted at the various stakeholders involved in the process of prescribing, dispensing, and consumption of medicines. Many strategies aimed at modifying prescribing behavior have been proposed and adopted. The evidence to date indicates that mailed educational materials alone are not sufficient to modify prescribing behavior [173,180]. Early studies conducted in Australia [184] and Denmark [185] concluded that mailed, unsolicited, centralized, government-sponsored feedback, one based on aggregate prescribing data and the other with a clinical guideline, had no impact on physician prescribing. For interventions that have been shown to be effective in improving drug prescribing (discussed in Chapter 38), there is a need to further define their relative efficacy and proper role in a comprehensive strategy for optimizing drug utilization. Questions yet to be addressed through a proper methodology deal with the role of printed drug information such as drug bulletins; the duration

of effect of educational interventions such as group discussions, lectures, and seminars, each in both the outpatient as well as the inpatient settings; and the generalizability of face-to-face methods, as described by Avorn and Soumerai [173], Schaffner *et al.* [174], and Ray *et al.* [175]. There is also a need for more research on whether the benefits and savings achieved with intervention strategies outweighed the costs of performing the intervention.

More clinically applicable approaches to DUR programs, such as the computerized screening of patient-specific drug histories in outpatient care to prevent drug-induced hospitalizations, still require further development and assessment. Although numerous studies have described the results of these and other novel programs [177,178,186,187], adequate documentation of their efficacy in improving quality of care is an important subject for future work. Patient outcome measures as well as process measures of quality of drug utilization have to be included in such studies. To be effective and efficient, healthcare policy options should be based on sound scientific evidence [188].

Challenges

The use of computerized databases has greatly facilitated studies of drug utilization. Although useful, most of these databases are far from ideal, as they have been set up mainly for administrative purposes, such as reimbursement, and drug utilization data are obtained as "spin-off" information. The model information system that will suit both medical and administrative needs [189] is still unavailable, although there is increasing use of electronic medical records for routine practice in countries such as the Netherlands, Australia, the UK, and the US. There is also a general lack of patient-level databases on medicine use in inpatient care [190]. Existing medical and pharmaceutical databases, with all their described limitations, will continue to be the main resources for these drug utilization studies. There is, however, a rapid growth of data coming from other sources. Even though new computer techniques and machine learning have been developed, many challenges remain, such as how to deal with missing data, unstructured data, poor data validity, and interoperability.

We have been fortunate to live in an era when large amounts of drug utilization data have become available for research. It is important to recognize, however, that the increasing amounts of digitalized personal data may add fuel to the debate on confidentiality. Confidentiality of patient data has so far been successfully handled and procedures have been implemented in most countries that are consistent with the guidelines for good practice in data privacy, medical record confidentiality, and research developed by the International Society for Pharmacoepidemiology (ISPE). Still, it is important to acknowledge the ongoing debate. In the EU, a new data protection framework is being implemented in all member states. Similar initiatives are being taken in other countries. Hopefully, this will not prevent opportunities for conducting research.

Even though drug utilization analyses today are conducted routinely in most health systems, this does not imply that drug utilization research is awarded high priority. The recruitment and training of researchers may be hampered by limitations in funding, as well as limitations in career opportunities. These two problems impose constraints on the future development of studies in drug utilization. However, despite this, the search continues for simple and relatively inexpensive methods to conduct descriptive studies of drug utilization, and effective intervention strategies that may contribute to the optimization of drug therapy. Fortunately, the increasing commitment to drug utilization research is reflected in the development and growth of international groups such as ISPE (www.pharmacoepi.org) [191], the International Clinical Epidemiology Network (INCLEN; www.inclentrust.org) [192], EuroDURG (www.

eurodurg.com) [193], DURG-LA [28], and INRUD (www.msh.org/INRUD) [194,195].

In summary, the study of drug utilization continues to evolve. The development of computerized databases is allowing the linkage of drug utilization data to clinical data and much other information to get a better understanding of drug utilization. The WHO/INRUD indicator-based approach to drug utilization studies is facilitating the development of drug utilization research in developing and transitional countries. Many strategies have already been proposed, tested, and implemented to improve the quality of drug prescribing in developed [196] and developing countries [197]. DUR programs, particularly approaches that take into primary consideration patient outcome measures, merit further rigorous study and improvement. Opportunities for the study of drug utilization are still underexplored, but the political issue regarding the confidentiality of medical records, as well as limitations in funding and personnel, may limit the growth of drug utilization research.

References

1 Bergman U, Grímsson A, Wahba AHW, Westerholm BE. *Studies in Drug Utilization. European Series No. 8*. Copenhagen: World Health Organization, Regional Office for Europe, 1979.

2 Dukes MNG ed. *Drug Utilization Studies: Methods and Uses. European Series No. 8*. Copenhagen: World Health Organization, Regional Office for Europe, 1993.

3 Ferrer P, Ballarin E, Sabate M, *et al*. Sources of European drug consumption data at a country level. *Int J Public Health* 2014; **59**(5): 877–87.

4 Sabaté M, Ferrer P, Ballarín E, *et al*. Inpatient drug utilization in Europe: nationwide data sources and a review of publications on a selected group of medicines (PROTECT Project). *Basic Clin Pharmacol Toxicol* 2015; **116**(3): 201–11.

5 Bergman U, Elmes P, Halse M, *et al*. The measurement of drug consumption: drugs for diabetes in Northern Ireland, Norway and Sweden. *Eur J Clin Pharmacol* 1975; **8**(2): 83–9.

6 Sjöqvist F, Agenäs I. Drug utilization studies: implications for medical care. In Sjöqvist F, Agenäs I, ed. *Proceedings from ANIS Symposium, Sånga-Säby, Sweden, June 8–9, 1982*. Stockholm: Apoteksbolaget, 1984.

7 Bergman U, Andersen M, Vaccheri A, Bjerrum L, Wettermark B, Montanaro N. Deviations from evidence-based prescribing of non-steroidal anti-inflammatory drugs in three European regions. *Eur J Clin Pharmacol* 2000; **56**(3): 269–72.

8 Vlahović-Palčevski V, Wettermark B, Bergman U. Quality of non-steroidal anti-inflammatory drug prescribing in Croatia (Rijeka) and Sweden (Stockholm). *Eur J Clin Pharmacol* 2002; **58**(3): 209–14.

9 Cars O, Mölstad S, Melander A. Variation in antibiotic use in the European Union. *Lancet* 2001; **357**(9271): 1851–3.

10 Magrini N, Einarson T, Vaccheri A, McManus P, Montanaro N, Bergman U. Use of lipid-lowering drugs from 1990 to 1994: an international comparison among Australia, Finland, Italy (Emilia Romagna Region), Norway and Sweden. *Eur J Clin Pharmacol* 1997; **53**(3): 185–9.

11 Griffiths K, McDevitt DG, Andrew M, *et al*. Validation of observed differences in the utilization of antihypertensive and antidiabetic drugs in Northern Ireland, Norway and Sweden. *Eur J Clin Pharmacol* 1985; **29**(1): 1–8.

12 Griffiths K, McDevitt DG, Andrew M, *et al*. Therapeutic traditions in Northern Ireland, Norway and Sweden: II. Hypertension. *Eur J Clin Pharmacol* 1986; **30**(5): 521–5.

13 Vancheri F, Backlund L, Strender LE, Godman B, Wettermark B. Time trends in statin utilisation and coronary mortality in Western

European countries. *BMJ Open* 2016; **6**(3): e010500.

14 Stålhammar, J, Bergman, U, Boman, K, Dahlen M. Metabolic control in diabetic subjects in three Swedish areas with high, medium, and low sales of antidiabetic drugs. *Diabetes Care* 1991; **14**: 12–19.

15 Baum C, Kennedy DL, Forbes MB, Jones JK. Drug use and expenditures in 1982. *JAMA* 1985; **253**(3): 382–6.

16 Stolley PD, Lasagna L. Prescribing patterns of physicians. *J Chronic Dis* 1969; **22**(6–7): 395–405.

17 Scheckler WE, Bennett JV. Antibiotic usage in seven community hospitals. *JAMA* 1970; **213**(2): 264–7.

18 Castle M, Wilfret CM, Cate T, *et al*. Antibiotic use at Duke University Medical Center. *JAMA* 1977; **237**(26): 2819–22.

19 Schollenberg E, Albritton WL. Antibiotic misuse in a pediatric teaching hospital. *Can Med Assoc J* 1980; **122**(1): 49–52.

20 Soumerai SB, McLaughlin TJ, Spiegelman D, Hertzmark E, Thibault G, Goldman L. Adverse outcomes of underuse of beta-blockers in elderly survivors of acute myocardial infarction. *JAMA* 1997; **277**(2): 115–21.

21 Antani MR, Beyth RJ, Covinsky KE, *et al*. Failure to prescribe warfarin to patients with nonrheumatic atrial fibrillation. *J Gen Intern Med* 1996; **11**(12): 713–20.

22 McLaughlin TJ, Soumerai SB, Willison DJ, *et al*. Adherence to national guidelines for drug treatment of suspected acute myocardial infarction: evidence for undertreatment in women and the elderly. *Arch Intern Med* 1996; **156**(7): 799–805.

23 O'Connor GT, Quinton HB, Traven ND, *et al*. Geographic variation in the treatment of acute myocardial infarction: the cooperative cardiovascular project. *JAMA* 1999; **281**(7): 627–33.

24 Kaufman DW, Kelly JP, Rosenberg L, Anderson TE, Mitchell AA. Recent patterns of medication use in the ambulatory adult population of the United States: the Slone survey. *JAMA* 2002; **287**(3): 337–44.

25 Kelly JP, Kaufman DW, Kelley K, Rosenberg L, Anderson TE, Mitchell AA. Recent trends in use of herbal and other natural products. *Arch Intern Med* 2005; **165**(3): 281–6.

26 Vernacchio L, Kelly JP, Kaufman DW, Mitchell AA. Cough and cold medication use by US children, 1999–2006: results from the Slone survey. *Pediatrics* 2008; **122**(2): e323–e329.

27 Vernacchio L, Kelly JP, Kaufman DW, Mitchell A. Pseudoephedrine use among U.S. children 1999–2006: results from the Slone survey. *Pediatrics* 2008; **122**: 1299–304.

28 Buschiazzo H, Chaves A, Figueras A, Laporte J-R. Drug utilization in Latin America: the example of DURG-LA. *WHO Essent Drugs Monit* 2003; **32**. http://apps.who.int/medicinedocs/en/d/Js4940e/12.html (accessed May 2019).

29 Durán CE, Christiaens T, Acosta Á, Vander Stichele R. Systematic review of cross-national drug utilization studies in Latin America: methods and comparability. *Pharmacoepidemiol Drug Saf* 2016; **25**(1): 16–25.

30 Hogerzeil H. Trends in drug utilization research in developing countries. *WHO Drug Inf* 2002; **16**. http://apps.who.int/medicinedocs/en/d/Js4951e/5.3.html (accessed May 2019).

31 AsPEN collaborators; Andersen M, Bergman U, Choi NK, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; **22**(7): 700–4.

32 Elseviers M, Wettermark B, Almarsdottir AB, *et al*., eds. *Drug Utilization Research Methods and Applications*. Chichester: John Wiley & Sons, 2016.

33 WHO Expert Committee. The selection of essential drugs. Technical Report Series No. 615. Geneva: World Health Organization, 1977.

34 Lunde PKM, Baksaas I. Epidemiology of drug utilization: basic concepts and methodology. *Acta Med Scand* 1987; **222**(721 S): 7–11.

35 Brodie DC. Drug utilization and drug utilization review and control. NCHS-RD-70-8. Rockville, MD: Department of

Health, Education, and Welfare, National Center for Health Services Research and Development, Health Services and Mental Health Administration, 1970.
36 Conley BE. *Social and Economic Aspects of Drug Utilization Research*. Hamilton, IL: Drug Intelligence Publications, 1976.
37 Academy of Managed Care Pharmacy. Drug utilization review. 2009. www.amcp.org/WorkArea/DownloadAsset.aspx?id=9296 (accessed January 2018).
38 Laporte J, Porta M, Capella D. Drug utilization studies: a tool for determining the effectiveness of drug use. *Br J Clin Pharmacol* 1983; **16**(3): 301–4.
39 Lunde PKM, Baksaas I. Epidemiology of drug utilization: basic concepts and methodology. *Acta Med Scand Suppl* 1988; **721**: 7–11.
40 Crooks J. Drug epidemiology and clinical pharmacology: their contributions to patient care. *Br J Clin Pharmacol* 1983; **16**: 351–7.
41 Hemminki E. Review of literature on the factors affecting drug prescribing. *Soc Sci Med* 1975; **9**: 111–15.
42 Christensen DB, Bush PJ. Drug prescribing: patterns, problems and proposals. *Soc Sci Med Part A Med Psychol Med Sociol* 1981; **15**(3): 343–55.
43 Harbarth S, Monnet D. Cultural and socioeconomic determinants of antibiotic use. In Gould IM, Van Der Meer J, ed. *Antibiotic Policies: Fighting Resistance*. New York: Springer, 2007, pp. 29–40.
44 Avorn J, Soumerai SB. Use of a computer-based Medicaid drug data to analyze and correct inappropriate medication use. *J Med Syst* 1982; **6**(4): 377–86. http://link.springer.com/content/pdf/10.1007/BF00992880 (accessed May 2019).
45 Hofstede GH. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*, 2nd edn. Thousand Oaks, CA: Sage.
46 Reginald D, Grigoryan L, Stålsby Lundborg C, *et al*. Are cultural dimensions relevant for explaining cross-national differences in antibiotic use in Europe? *BMC Health Serv Res* 2008; **8**(1): 123.
47 Hulscher ME, Grol RP, van der Meer JW. Antibiotic prescribing in hospitals: a social and behavioural scientific approach. *Lancet Infect Dis* 2010; **10**(3): 167–75.
48 Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989; **67**(2): 268–317.
49 Anderson GM, Lexchin J. Strategies for improving prescribing practice. *CMAJ* 1996; **154**(7): 1013–17.
50 Farmer AP, Légaré F, Turcot L, *et al*. Printed educational materials: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2008; **3**: CD004398.
51 Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; **342**(8883): 1317–22.
52 Sjöqvist F, Birkett D. Drug utilization. *Introduction to Drug Utilization Research*. World Health Organization, 2003, pp. 76–84. https://apps.who.int/medicinedocs/en/d/Js4876e/1.html (accessed May 2019).
53 Melmon K. Preventable drug reactions: causes and cures. *N Engl J Med* 1971; **284**(24): 1361–8.
54 Pirmohamed M, James S, Meakin S, *et al*. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 2004; **329**(7456): 15–19.
55 Impicciatore P, Choonara I, Clarkson A, Provasi D, Pandolfini C, Bonati M. Incidence of adverse drug reactions in paediatric in/out-patients: a systematic review and meta-analysis of prospective studies. *Br J Clin Pharmacol* 2001; **52**(1): 77–83.
56 Runciman WB, Roughead EE, Semple SJ, Adams RJ. Adverse drug events and medication errors in Australia. *Int J Qual Heal Care* 2003; **15**(suppl 1): i49–i59.
57 Samoy LJ, Zed PJ, Wilbur K, Balen RM, Abu-Laban RB, Roberts M. Drug-related hospitalizations in a tertiary care internal

medicine service of a Canadian hospital: a prospective study. *Pharmacotherapy* 2006; **26**(11): 1578–86.

58 Von Euler M, Eliasson E, Hlén G, Bergman U. Adverse drug reactions causing hospitalization can be monitored from computerized medical records and thereby indicate the quality of drug utilization. *Pharmacoepidemiol Drug Saf* 2006; **15**: 179–84.

59 Bergman U. *Focusrapport. Läkemedelsbiverkningar Som Orsak till Inläggning På Sjukhus (Adverse Drug Reactions (ADRs) Causing Hospitalizations: Review of Swedish Studies)*. Stockholm: Läns Landsting, 2005.

60 Matanović SM, Vlahović-Palčevski V. Potentially inappropriate medications in the elderly: a comprehensive protocol. *Eur J Clin Pharmacol* 2012; **68**(8): 1123–38.

61 Matanović SM, Vlahović-Palčevski V. Potentially inappropriate prescribing to the elderly: comparison of new protocol to Beers criteria with relation to hospitalizations for ADRs. *Eur J Clin Pharmacol* 2014; **70**(4): 483–90.

62 Bates DW. Medication errors: how common are they and what can be done to prevent them? *Drug Saf* 1996; **15**(5): 303–10.

63 Haynes RB. Introduction. In Haynes BR, Taylor DW, Sackett DL, eds., *Compliance in Health Care*. Baltimore, MD: Johns Hopkins University Press, 1979, pp. 5–16.

64 Andrade SE, Walker AM, Gottlieb LK, *et al*. Discontinuation of antihyperlipidemic drugs: do rates reported in clinical trials reflect rates in primary care settings? *N Engl J Med* 1995; **332**(17): 1125–31.

65 Avorn J, Monette J, Lacour A, *et al*. Persistence of use of lipid-lowering medications: a cross-national study. *JAMA* 1998; **279**(18): 1458–62.

66 Larsen J, Vaccheri A, Andersen M, Montanaro N, Bergman U. Lack of adherence to lipid-lowering drug treatment: a comparison of utilization patterns in defined populations in Funen, Denmark and Bologna, Italy. *Br J Clin Pharmacol* 2000; **49**(5): 463–71.

67 Isacsson G, Boëthius G, Bergman U. Low level of antidepressant prescription for people who later commit suicide: 15 years of experience from a population-based drug database in Sweden. *Acta Psychiatr Scand* 1992; **85**(6): 444–8.

68 Whittle J, Wickenheiser L, Venditti LN. Is warfarin underused in the treatment of elderly persons with atrial fibrillation? *Arch Intern Med* 1997; **157**(4): 441–5.

69 Albers GW, Yim JM, Belew KM, *et al*. Status of antithrombotic therapy for patients with atrial fibrillation in university hospitals. *Arch Intern Med* 1996; **156**(20): 2311–16.

70 Mendelson, G, Aronow W. Underutilization of measurement of serum low-density lipoprotein cholesterol levels and of lipid-lowering therapy in older patients with manifest atherosclerotic disease. *J Am Geriatr Soc* 1998; **46**(9): 1128–31.

71 Aronow W. Underutilization of lipid-lowering drugs in older persons with prior myocardial infarction and a serum low-density lipoprotein cholesterol > 125 mg/dl. *Am J Cardiol* 1998; **82**(5): 668–9.

72 Austin PC, Mamdani MM, Juurlink DN, Alter DA, Tu JV. Missed opportunities in the secondary prevention of myocardial infarction: an assessment of the effects of statin underprescribing on mortality. *Am Heart J* 2006; **151**(5): 969–75.

73 Krumholz HM, Radford MJ, Ellerbeck EF, *et al*. Aspirin in the treatment of acute myocardial infarction in elderly Medicare beneficiaries: patterns of use and outcomes. *Circulation* 2006; **114**(10): 2841–7.

74 Wirtz VJ, Kaplan WA, Kwan GF, Laing RO. Access to medications for cardiovascular diseases in low- and middle-income countries. *Circulation* 2016; **133**(21): 2076–85.

75 Amstislavski P, Matthews A, Sheffield S, Maroko AR, Weedon J. Medication deserts:

survey of neighborhood disparities in availability of prescription medications. *Int J Health Geogr* 2012; **11**: 48.

76 Berkowitz SA, Seligman HK, Choudhry NK. Treat or eat: food insecurity, cost-related medication underuse, and unmet needs. *Am J Med* 2014; **127**(4): 303–10.

77 Haider BA, Lassi ZS, Bhutta ZA. Short-course versus long-course antibiotic therapy for non-severe community-acquired pneumonia in children aged 2 months to 59 months. *Cochrane Database Syst Rev* 2008; **2**: CD005976.

78 Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med* 1985; **78**(3): 475–80.

79 Lasagna L. A plea for the "naturalistic" study of medicines. *Eur J Clin Pharmacol* 1974; **7**(3): 153–4.

80 Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ* 1995; **311**(6996): 42–5.

81 Almarsdottir AB, Bastholm Rahmner P. Qualitative methods in drug utilization. In Elseviers M, Wettermark B, Almarsdottir AB, *et al*., eds., *Drug Utilization Research. Methods and Applications*. Chichester: John Wiley & Sons, 2016, pp. 139–49.

82 Lasagna L. Society overmedicated? *Lancet* 1974; **304**(7883): 781.

83 Qvarnström M, Kahan T, Kieler H, *et al*. Persistence to antihypertensive drug treatment in Swedish primary healthcare. *Eur J Clin Pharmacol* 2013; **69**(11): 1955–64.

84 Alfian SD, Worawutputtapong P, Schuiling-Veninga CC, *et al*. Pharmacy-based predictors of non-persistence and non-adherence to statin treatment among patients on oral diabetes medication in the Netherlands. *Curr Med Res Opin* 2018; **34**(6):1013–19.

85 Forslund T, Wettermark B, Hjemdahl P. Comparison of treatment persistence with different oral anticoagulants in patients with atrial fibrillation. *Eur J Clin Pharmacol* 2016; **72**(3): 329–38.

86 Gusmão R, Quintão S, McDaid D, *et al*. Antidepressant utilization and suicide in Europe: an ecological multi-national study. *PLoS One* 2013; **8**(6): e66455.

87 Sofianopoulou E, Rushton SP, Diggle PJ, Pless-Mulloli T. Association between respiratory prescribing, air pollution and deprivation, in primary health care. *J Public Health (Oxf)* 2013; **35**(4): 502–9.

88 Vancheri F, Wettermark B, Strender LE, Godman B, Wettermark B. Trends in coronary heart disease mortality and statin utilization in two European areas with different population risk levels: Stockholm and Sicily. *Int Cardiovasc Forum J* 2014; **3**: 140–6.

89 Kozman D, Graziul C, Gibbons R, Alexander GC. Association between unemployment rates and prescription drug utilization in the United States, 2007–2010. *BMC Health Serv Res* 2012; **12**(1): 435.

90 Rucker T. Drug use: data, sources, and limitations. *JAMA* 1974; **230**(6): 888–90.

91 *Consumption of Drugs: Report on a Symposium in Oslo, 1969*. EURO 3102. Copenhagen: WHO Regional Office for Europe, 1970.

92 Hallas J, Støvring H. Templates for analysis of individual-level prescription data. *Basic Clin Pharmacol Toxicol* 2006; **98**(3): 260–5.

93 Bergman U. Pharmacoepidemiology – from description to quality assessment: a Swedish perspective. *Nor J Epidemiol* 2001; **11**(1): 31–6.

94 Strom BL, Carson JL, Halpern AC, *et al*. Using a claims database to investigate drug-induced Stevens-Johnson syndrome. *Stat Med* 1991; **10**(4): 565–76.

95 Kothari R, Gaurav Garg C. *Research Methodology: Methods and Techniques*, 3rd edn. New Delhi: New Age International, 2014.

96 Donnan P, Steinke D, Newton R, Morris A. Changes in treatment after the start of oral hypoglycaemic therapy in Type 2 diabetes: a population based study. *Diabet Med* 2002; **19**(1): 606–10.

97 Eriksson I, Ibanez L. Secondary data sources for drug utilization research. In Elseviers M,

Wettermark B, Almarsdottir AB, *et al*., eds., *Drug Utilization Research: Methods and Applications*. Chichester: John Wiley & Sons, 2016, pp. 39–48.

98 Hallas J. Conducting pharmacoepidemiologic research in Denmark. *Pharmacoepidemiol Drug Saf* 2001; **10**(7): 619–23.

99 Vaccheri A, Bjerrum L, Resi D, Bergman U, Montanaro N. Antibiotic prescribing in general practice: striking differences between Italy (Ravenna) and Denmark (Funen). *J Antimicrob Chemother* 2002; **50**(6): 989–97.

100 Wettermark B, Hammar N, Fored CM, *et al*. The new Swedish Prescribed Drug Register: opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007; **16**(7): 726–35.

101 Wallerstedt SM, Wettermark B, Hoffmann M. The first decade with the Swedish Prescribed Drug Register: a systematic review of the output in the scientific literature. *Basic Clin Pharmacol Toxicol* 2016; **119**(5): 464–9.

102 Williams D, Feely J. Pharmacoepidemiology: an Irish perspective. *Pharmacoepidemiol Drug Saf* 2001; **10**(7): 641–5.

103 Hux JE, Kopp A, Mamdani MM. Turning 65 in Ontario: the impact of public drug benefit coverage on hospitalizations for acute and chronic disease. *Healthc Policy* 2006; **1**(3): 87–98.

104 Crystal S, Akincigil A, Bilder S, Walkup JT. Studying prescription drug use and outcomes with medicaid claims data: strengths, limitations, and strategies. *Med Care* 2007; **45**(10 Supl 2): S58–S65.

105 Kimura T, Matsushita Y, Yang YHK, Choi NK, Park BJ. Pharmacovigilance systems and databases in Korea, Japan, and Taiwan. *Pharmacoepidemiol Drug Saf* 2011; **20**(12): 1237–45.

106 Bromley S, De Vries C, Farmer R. Utilisation of hormone replacement therapy in the United Kingdom: a descriptive study using the general practice research database. *BJOG* 2004; **111**(4): 369–76.

107 Nash D. National Drug and Therapeutic Index. *P T* 2002; **27**: 530. https://www.ptcommunity.com/journal/article/archives/2002/11/530/national-drug-and-therapeutic-index (accessed May 2019).

108 Stolar M. Drug use review: operational definitions. *Am J Hosp Pharm* 1978; **35**: 76–8.

109 Todd M. Drug use evaluation. In Brown, TR; American Society of Hospital Pharmacists, eds., *Handbook of Institutional Pharmacy Practice*, 3rd edn. Bethesda, MD: American Society of Hospital Pharmacists, 1992, pp. 261–71.

110 Baksaas I, Lunde PK. National drug policies: the need for drug utilization studies. *Trends Pharmacol Sci* 1986; **7**(1): 331–4.

111 Westerholm B. Therapeutic auditing at the national and international level. *Br J Clin Pharmacol* 1986; **22**(Suppl 1): 55S–59S.

112 *Guidelines for ATC Classification and DDD Assignment 2018*. Oslo: WHO Collaborating Centre for Drug Statistics Methodology, 2017.

113 Lee D, Chaves Matamoros A, Mora Duarte J. Changing patterns of antibiotic utilization patterns in Costa Rica. *APUA Newslett* 1991; **9**: 7.

114 Nappo S, Carlini EA. Preliminary finding: consumption of benzodiazepines in Brazil during the years 1988 and 1989. *Drug Alcohol Depend* 1993; **33**(1): 11–17.

115 Truter I, Wiseman IC, Kotze TJ. The defined daily dose as a measure of drug consumption in South Africa: a preliminary study. *South African Med J* 1996; **86**(6): 675–9.

116 Vlahović-Palčevski V, Palcevski G, Mavric Z, Francetic I. Factors influencing antimicrobial utilization at a university hospital during a period of 11 years. *Int J Clin Pharmacol Ther* 2003; **41**(7): 287–93.

117 Ansari F. Use of systemic anti-infective agents in Iran during 1997–1998. *Eur J Clin Pharmacol* 2001; **57**(6): 547–51.

118 Mond J, Morice R, Owen C, Korten A. Use of antipsychotic medications in Australia between July 1995 and December 2001. *Aust N Z J Psychiatry* 2003; **37**(1): 55–61.

119 WHO International Working Group for Drug Statistics Methodology. Introduction to drug utilization research. *Solutions* 2003; 1–48.

120 Bergman U, Boman G, Wiholm BE. Epidemiology of adverse drug reactions to phenformin and metformin. *BMJ* 1978; **2**(6135): 464–66.

121 Schüssel K, Schulz M. Prescribing of COX-2 inhibitors in Germany after safety warnings and market withdrawals. *Pharmazie* 2006; **61**(10): 878–86.

122 Wirtz VJ, Dreser A, Gonzales R. Trends in antibiotic utilization in eight Latin American countries, 1997–2007. *Rev Panam Salud Pública* 2010; **27**(3): 219–25.

123 Goossens H, Ferech M, Vander Stichele R, Elseviers M. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 2005; **365**(9459): 579–87.

124 Bronzwaer SLAM, Cars O, Buchholz U, *et al*. The relationship between antimicrobial use and antimicrobial resistance in Europe. *Emerg Infect Dis* 2002; **8**(3): 278–82.

125 Gulbinovic J, Myrbäck KE, Bytautienė J, Wettermark B, Struwe J, Bergman U. Marked differences in antibiotic use and resistance between university hospitals in Vilnius, Lithuania, and Huddinge, Sweden. *Microb Drug Resist* 2001; **7**(4): 383–9.

126 Mölstad S, Erntell M, Hanberger H, *et al*. Sustained reduction of antibiotic use and low bacterial resistance: 10-year follow-up of the Swedish Strama programme. *Lancet Infect Dis* 2008; **8**(2): 125–32.

127 European Centre for Disease Prevention and Control. About the network. https://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/esac-net-about (accessed May 2019).

128 Stanulovic M, Milosev M, Jakovljevic V, Roncevic N. Epidemiological evaluation of anti-infective drug prescribing for children in outpatient practice. *Dev Pharmacol Ther* 1987; **10**: 278–91.

129 Bergman U, Sjöqvist F. Measurement of drug utilization in Sweden: methodological and clinical implications. *Acta Med Scand* 1984; **105**(Suppl 683): 15–22.

130 Bergman U. The history of the Drug Utilization Research Group in Europe. *Pharmacoepidemiol Drug Saf* 2006; **15**(2): 95–8.

131 Bergman U, Dahlström M. Benzodiazepine utilization in Sweden and other Nordic countries. In Lader, MH, Davies H, eds., *Drug Treatment of Neurotic Disorders*. Edinburgh: Churchill Livingstone, 1986, pp. 43–52.

132 Isacsson G, Holmgren P, Druid H, Bergman U. The utilization of antidepressants – a key issue in the prevention of suicide: an analysis of 5281 suicides in Sweden during the period 1992–1994. *Acta Psychiatr Scand* 1997; **96**(2): 94–100.

133 Barlam TF, Cosgrove SE, Abbo LM, *et al*. Executive summary: implementing an antibiotic stewardship program: guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Clin Infect Dis* 2016; **62**(10): 1197–202.

134 Rønning M, Blix HS, Harbø BT, Strøm HS. Different versions of the anatomical therapeutic chemical classification system and the defined daily dose: are drug utilisation data comparable? *Eur J Clin Pharmacol* 2000; **56**(9–10): 723–7.

135 Caja Costarricense de Seguro Social. *Lista Oficial Medicamentos*. Costa Rica: CCSS, 2017. https://www.ccss.sa.cr/lom (accessed May 2019).

136 Republic of South Africa. *Essential Drugs Programme. Primary Healthcare. Standard Treatment Guidelines and Essential Medicines List*, 5th edn. Pretoria: National Department of Health, 2014.

137 Ministry of Health Malaysia. Malaysian Statistics on Medicines 2009 & 2010. 2014.

http://www.pharmacy.gov.my/v2/sites/default/files/document-upload/malaysian-statistics-medicines-2009-2010_0.pdf (accessed May 2019).

138 Pahor M, Chrischilles EA, Guralnik JM, Brown SL, Wallace RB, Carbonin P. Drug data coding and analysis in epidemiologic studies. *Eur J Epidemiol* 1994; **10**(4): 405–11.

139 US Department of Veterans Affairs. VA National Formulary. https://www.pbm.va.gov/nationalformulary.asp (accessed January 2018).

140 BMJ Group and Royal Pharmaceutical Society. *British National Formulary, BNF 74*. London: BMJ Group and Pharmaceutical Press, 2017.

141 World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*, 10th Revision. Geneva: WHO, 2016.

142 Campbell S, Wettermark B, Andersen M. Defining and developing quality indicators for drug utilization. In Elseviers M, Wettermark BAA, Andersen M, *et al*., eds., *Drug Utilization Research: Methods and Uses*. Chichester: Wiley Blackwell, 2016, pp. 126–39.

143 Muijrers PEM, Janknegt R, Sijbrandij J, Grol RPTM, Knottnerus JA. Prescribing indicators: development and validation of guideline-based prescribing indicators as an instrument to measure the variation in the prescribing behaviour of general practitioners. *Eur J Clin Pharmacol* 2004; **60**(10): 739–46.

144 Fernández Urrusuno R, Montero Balosa MC, Pérez Pérez P, Pascual De La Pisa B. Compliance with quality prescribing indicators in terms of their relationship to financial incentives. *Eur J Clin Pharmacol* 2013; **69**(10): 1845–53.

145 Godman B, Shrank W, Andersen M, *et al*. Policies to enhance prescribing efficiency in Europe: findings and future implications. *Front Pharmacol* 2011; **1**: 41.

146 Roughead EE, Zhang F, Ross-Degnan D, Soumerai S. Differential effect of early or late implementation of prior authorization policies on the use of Cox II inhibitors. *Med Care* 2006; **44**(4): 378–82.

147 Beers MH. Explicit criteria for determining potentially inappropriate medication use by the elderly: an update. *Arch Intern Med* 1997; **157**: 1531–6.

148 Samuel MJ. American Geriatrics Society 2015 updated Beers criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc* 2015; **63**(11): 2227–46.

149 Carnovale C, Conti V, Perrone V, *et al*. Paediatric drug use with focus on off-label prescriptions in Lombardy and implications for therapeutic approaches. *Eur J Pediatr* 2013; **172**(12): 1679–85.

150 Popović B, Quadranti NR, Matanović SM, *et al*. Potentially inappropriate prescribing in elderly outpatients in Croatia. *Eur J Clin Pharmacol* 2014; **70**(6): 737–44.

151 Wettermark B, Pehrsson A, Jinnerot D, Bergman U. Drug utilisation 90% profiles: a useful tool for quality assessment of prescribing in primary health care in Stockholm. *Pharmacoepidemiol Drug Saf* 2003; **12**(6): 499–510.

152 Gustafsson LL, Wettermark B, Godman B, *et al*. The "Wise List": a comprehensive concept to select, communicate and achieve adherence to recommendations of essential drugs in ambulatory care in Stockholm. *Basic Clin Pharmacol Toxicol* 2011; **108**(4): 224–33.

153 Teichert M, Van Der Aalst A, De Wit H, Stroo M, De Smet PAGM. How useful are prescribing indicators based on the DU90% method to distinguish the quality of prescribing between pharmacotherapy audit meetings with different levels of functioning? *Eur J Clin Pharmacol* 2007; **63**(12): 1171–7.

154 Krivoy N, El-Ahal WA, Bar-Lavie Y, Haddad S. Antibiotic prescription and cost patterns in a general intensive care unit. *Pharm Pract (Granada)* 2007; **5**(2): 67–73.

155 Wettermark B, Pehrsson A, Juhasz-Haverinen M, *et al*. Financial incentives linked to self-assessment of prescribing

patterns: a new approach for quality improvement of drug prescribing in primary care. *Qual Prim Care* 2009; **17**(3): 179–89.

156 Eriksen J, Gustafsson LL, Ateva K, *et al*. High adherence to the "Wise List" treatment recommendations in Stockholm: a 15-year retrospective review of a multifaceted approach promoting rational use of medicines. *BMJ Open* 2017; **7**(4): e014345.

157 Goryachkina K, Babak S, Burbello A, Wettemark B, Bergman U. Quality use of medicines: a new method of combining antibiotic consumption and sensitivity data – application in a Russian hospital. *Pharmacoepidemiol Drug Saf* 2008; **17**(6): 636–44.

158 Linnarsson R. Drug interactions in primary health care: a retrospective database study and its implications for the design of a computerized decision support system. *Scand J Prim Heal Care* 1993; **11**: 181–6.

159 Palčevski G, Skočibušić N, Vlahović-Palčevski V. Unlicensed and off-label drug use in hospitalized children in Croatia: a cross-sectional survey. *Eur J Clin Pharmacol* 2012; **68**(7): 1073–7.

160 Arnlind MH, Wettermark B, Nokela M, Hjemdahl P, Rehnberg C, Jonsson EW. Regional variation and adherence to guidelines for drug treatment of asthma. *Eur J Clin Pharmacol* 2010; **66**(2): 187–98.

161 Frisk P, Mellgren TO, Hedberg N, Berlin A, Granath F, Wettermark B. Utilisation of angiotensin receptor blockers in Sweden: combining survey and register data to study adherence to prescribing guidelines. *Eur J Clin Pharmacol* 2008; **64**(12): 1223–9.

162 Chen TJ, Chou LF, Hwang S. Trends in prescribing proton pump inhibitors in Taiwan: 1997–2000. *Int J Clin Pharmacol Ther* 2003; **41**(5): 207–12.

163 Pelaez-Ballestas I, Hernandez-Garduno A, Arredondo-Garcia JL, Viramontes-Madrid JL, Aguilar-Chiu A. Use of antibiotics in upper respiratory infections on patients under 16 years old in private ambulatory medicine. *Salud Publica Mex* 2003; **45**(3): 159–64.

164 Kahan E, Kahan NR, Chinitz DP. Urinary tract infection in women – physician's preferences for treatment and adherence to guidelines: a national drug utilization study in a managed care setting. *Eur J Clin Pharmacol* 2003; **59**(8–9): 663–8.

165 World Health Organization. *How to Investigate Drug Use in Health Facilities: Selected Drug Use Indicators*. Geneva: WHO, 1993.

166 Trap B. Prescription habits of dispensing and non-dispensing doctors in Zimbabwe. *Health Policy Plan* 2002; **17**(3): 288–95.

167 Santaso B. From research to action: the Gunungkidul experience. *Essent Drugs Monit* 1995; **2**: 21–2.

168 Santoso B. Small group intervention vs formal seminar for improving appropriate drug use. *Soc Sci Med* 1996; **42**(8): 1163–8.

169 Ofori-Adjei D, Arhinful DK. Effect of training on the clinical management of malaria by medical assistants in Ghana. *Soc Sci Med* 1996; **42**(8): 1169–76.

170 Hadiyono JE, Suryawati S, Danu SS, Sunartono, Santoso B. Interactional group discussion: results of a controlled trial using a behavioral intervention to reduce the use of injections in public health facilities. *Soc Sci Med* 1996; **42**(8): 1177–83.

171 World Health Organization. *Medicines Use in Primary Care in Developing and Transitional Countries: Fact Book Summarizing Results from Studies Reported between 1990 and 2006*. Geneva: WHO, 2009.

172 Chalker JC, Andualem T, Gitau LN, *et al*. Measuring adherence to antiretroviral treatment in resource-poor settings: the feasibility of collecting routine data for key indicators. *BMC Health Serv Res* 2010; **10**: 43.

173 Avorn J, Soumerai SB. Improving drug-therapy decisions through educational outreach: a randomized controlled trial of

academically based "detailing." *N Engl J Med* 1983; **308**(24): 1457–63.

174 Schaffner W, Ray WA, Federspiel CF, Miller WO. Improving antibiotic prescribing in office practice: a controlled trial of three educational methods. *JAMA* 1983; **250**(13): 1728–32.

175 Ray WA, Schaffner W, Federspiel CF. Persistence of improvement in antibiotic prescribing in office practice. *JAMA* 1985; **253**(12): 1774–6.

176 Strom BL, Morse ML. Use of computerized databases to survey drug utilization in relation to diagnoses. *Acta Med Scand* 1987; **222**(721 Suppl): 13–20.

177 Lee Morse M, Leroy AA, Gaylord TA, Kellenberger T. Reducing drug therapy-induced hospitalization: impact of drug utilization review. *Ther Innov Regul Sci* 1982; **16**(4): 199–202.

178 Groves R. Therapeutic drug-use review for the Florida Medicaid program 1985. *Am J Hosp Pharm* 1985; **42**: 316–19.

179 Hennessy S, Bilker WB, Zhou L, *et al*. Retrospective drug utilization review, prescribing errors, and clinical outcomes. *JAMA* 2003; **290**(11): 1494–9.

180 Soumerai, SB, Avorn J. Efficacy and cost-containment in hospital pharmacotherapy: state of the art and future directions. *Milbank Mem Fund Quart* 1984; **62**: 447–74.

181 Kidder D, Bae J. Evaluation results from prospective drug utilization review: Medicaid demonstrations. *Health Care Financ Rev* 1999; **20**(3): 107–18. http://www.ncbi.nlm.nih.gov/pubmed/10558015

182 EvaluatePharma. Surveying tomorrow's biopharma landscape. 2012. http://info.evaluatepharma.com/rs/evaluatepharmaltd/images/EvaluatePharma_NBI_Up_Close_2012.pdf (accessed May 2019).

183 Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment restrictions for prescription drugs under Medicaid: effects on therapy, cost, and equity. *N Engl J Med* 1987; **317**(9): 550–6.

184 O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in Australia. *BMJ* 1999; **318**(7182): 507–11.

185 Søndergaard J, Andersen M, Støvring H, Kragstrup J. Mailed prescriber feedback in addition to a clinical guideline has no impact: a randomised, controlled trial. *Scand J Prim Health Care* 2003; **21**(1): 47–51.

186 Monane M, Matthias DM, Nagle BA, Kelly MA. Improving prescribing patterns for the elderly through an online drug utilization review intervention: a system linking the physician, pharmacist, and computer. *JAMA* 1998; **280**(14): 1249–52.

187 Taylor L, Tamblyn R. Reasons for physician non-adherence to electronic drug alerts. *Stud Health Technol Inform* 2004; **107**: 1101–5.

188 Anderson GM, Spitzer WO, Weinstein MC, Wang E, Blackburn JL, Bergman U. Benefits, risks, and costs of prescription drugs: a scientific basis for evaluating policy options. *Clin Pharmacol Ther* 1990; **48**(2): 111–19.

189 Schiff GD, Rucker TD. Computerized prescribing: building the electronic infrastructure for better medication usage. *JAMA* 1998; **279**(13): 1024–9.

190 Larsen MD, Cars T, Hallas J. A minireview of the use of hospital-based databases in observational inpatient studies of drugs. *Basic Clin Pharmacol Toxicol* 2013; **112**(1): 13–18.

191 Lewis M. ISPE: present, past and future. *Pharmacoepidemiol Newsl* 1997; **11**: 4.

192 Guess HA, Goldsmith CH, Henry D, Strom BL. Pharmacoepidemiology as a focus for clinical epidemiology in developing countries. *J Clin Epidemiol* 1991; **45**(Suppl 2): 101–5.

193 Haaijer-Ruskamp, FM, Bergman U. Rational drug use in Europe – challenges for the 21st century. Report from the 1st meeting of EURO DURG, the European Drug

Utilization Research Group. *Eur J Clin Pharmacol* 1997; **52** (Suppl): I–VIII.

194 Ross-Degnan D, Laing R, Quick J, *et al*. A strategy for promoting improved pharmaceutical use: the International Network for Rational Use of Drugs. *Soc Sci Med* 1992; **35**(11): 1329–41.

195 Serradell J, Bjornson DC, Hartzema AG. Drug utilization study methodologies: national and international perspectives. *Drug Intell Clin Pharm* 1987; **21**(12): 994–1001.

196 Dartnell J. *Understanding, Influencing and Evaluating Drug Use*. Melbourne: Therapeutic Guidelines.

197 Rowe AK, Rowe SY, Vujicic M, *et al*. Review of strategies to improve health care provider performance. In Peters DH, El-Saharty S, Siadat B, Janovsky K, Vujicic M, eds., *Improving Health Service Delivery in Developing Countries: From Evidence to Action*. Washington, DC: World Bank, 2009, pp. 101–26.

19

Evaluating and Improving Physician Prescribing

Christine Y. Lu[1], the late Sumit R. Majumdar[2], Helene Lipton[3], and Stephen B. Soumerai[1]

[1] *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA*
[2] *Formerly Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada*
[3] *School of Medicine and Pharmacy, University of California at San Francisco, San Francisco, CA, USA*

> Research and clinical practice may be on parallel tracks headed in the same direction, but in contact only through rotting ties. [1]

The broad purposes of pharmacoepidemiology are to advance our knowledge of the risks and benefits of medication use in real-world populations, and to foster improved prescribing and patient health outcomes. If, however, physicians and other health practitioners fail to update their knowledge and practice in response to new and clinically important evidence on the outcomes of specific prescribing patterns, then the "fruits" of pharmacoepidemiologic research may have little impact on clinical practice.

It is for these reasons that a new discipline in the fields of health services research and clinical decision making has grown rapidly in importance – the science of assessing and improving clinical practices. The rapid growth of this new field (sometimes referred to as T-2 translational research or knowledge translation research) is based on the recognition that passive knowledge dissemination (e.g., publishing articles, distributing practice guidelines) is generally insufficient to improve clinical practices without supplemental behavioral change interventions based on relevant theories of diffusion of innovations, persuasive communications, and adult learning or social cognitive theory [1–10].

This chapter reviews some of these developments as they relate to medication use, defines several types of drug prescribing problems, discusses several thorny methodologic problems in this literature, reviews existing pharmacoepidemiologic and other evidence on the effectiveness of common interventions to improve prescribing, and concludes with a discussion of future research needs. For a more detailed and comprehensive examination of the literature on prescribing education, the role of the pharmacist as a change agent, disease management strategies for use in various settings, and the use of financial incentives and penalties, the reader is advised to consult several previous works

We dedicate this chapter to our colleague and friend Dr. Sumit Majumdar, whose untimely death occurred shortly after this chapter was written. We thank Caitlin Lupton MSc, of Harvard Pilgrim Health Care Institute for research assistance and administrative support.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

published elsewhere [11–33]. Portions of this chapter are derived from this body of work; in addition, we conducted computerized literature searches for papers published through early 2017, hand-searched our personal files and the cited references, and extensively consulted the Cochrane Effective Practice and Organisation of Care (EPOC) Group, a rigorous and continuously updated registry and synthesis of available evidence on studies of interventions to change physician behaviors [33].

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

There is little doubt that the importance of suboptimal prescribing practice (both underuse and overuse) vastly outweighs the costs of medications themselves [33–37] (see also Chapter 38). Drug therapies are the most common treatments in medical practice and more than three-quarters of all visits to a physician terminate with the writing of a prescription [15]; the potential for drug therapies for both alleviating and causing illness is illustrated throughout this book. As suggested by Lee [37], in this chapter we take a broad view of the concept of prescribing errors, and consider issues related to underuse, overuse, and misuse, since all contribute to the suboptimal utilization of pharmaceutical therapies. For example, we would consider as prescribing "errors" the following:

- Use of toxic or addictive drugs when safer and clinically appropriate agents are available (e.g., opioids instead of nonsteroidal anti-inflammatory drugs for pain [38,39]).
- Use of drug therapy when no therapy is required (e.g., antibiotics for viral respiratory infections).
- Use of an ineffective drug for a given indication (e.g., hormone therapy for prevention of cardiovascular disease in postmenopausal women).
- Use of a costly drug when a less expensive preparation would be just as effective (e.g., newer angiotensin-receptor blockers, instead of effective and inexpensive angiotensin-converting enzyme inhibitors or thiazide diuretics, for uncomplicated hypertension).
- Misuse of effective agents (e.g., too low doses of narcotic analgesics or too high dosages of benzodiazepines, when indicated, for the elderly).
- Failure to discontinue therapy when the drug is no longer needed (e.g., use of proton pump inhibitors for months to years in patients without documented gastroesophageal reflux disease).
- Failure to introduce new and effective drugs into practice (e.g., failure to use inhaled corticosteroids for asthma or spironolactone for heart failure).
- Failure to prescribe necessary drug therapies (e.g., failure to use beta-blockers following acute myocardial infarction or bisphosphonates after an osteoporotic fracture).
- Failure to achieve recommended therapeutic goals (e.g., failure to achieve systolic blood pressure levels below 130 mmHg or LDL cholesterol levels below 70 mg/dL for the secondary prevention of myocardial infarction).

Specific illustrations of these problem categories are ubiquitous in the literature. In the outpatient setting, numerous studies have documented that as much as 50% of antibiotic use is potentially inappropriate, with the unintended consequence that overuse of antibiotics may lead to the emergence of resistant pathogens [40]. A group at particular risk of iatrogenic injuries as a result of inappropriate medication exposure appears to be the frail elderly, whether they reside in the community or in nursing homes [34,41,42].

Because of the absence of diagnostic data in most published drug utilization research, and

because of the emphasis on cost containment within drug utilization review (DUR) programs (see subsequent definition), the existing literature may *underemphasize* the clinically important problem of underuse of highly effective medications. For example, Berlowitz *et al*. found that nearly 40% of patients with documented hypertension in the Veterans Administration (VA) healthcare system had uncontrolled hypertension (>160/90 mmHg), despite adequate healthcare and prescription drug coverage and more than six hypertension-related primary care visits each year [43]. Indeed, this demonstrates profound clinical inertia, as changes in antihypertensive therapy occurred in less than 10% of all of these visits [43]. In another study of 623 outpatients treated for acute myocardial infarction at the Yale-New Haven Hospital, researchers found that one-third of patients meeting strict randomized controlled trial (RCT) eligibility criteria for use of beta-blockers did not even receive a trial of therapy – contrary to existing guidelines. These patients experienced a 20–40% higher mortality rate postmyocardial infarction than may have been necessary [44]. There are many other examples of underuse and resultant unnecessary morbidity and mortality throughout the pharmacoepidemiologic literature.

Why do these problems occur? Can a comprehensive theory of behavioral change or knowledge translation provide the basis for programs designed to improve prescribing? Such an ideal model must be complex given the diversity of economic, organizational, educational, psychological, social, informational, and technological influences on daily prescribing practices [1–10,45–52]. Some of the factors responsible for suboptimal prescribing include the following:

- The failure of clinicians to keep abreast of important new findings on the risks and benefits of medications [6–8,45,52].
- Excessive promotion of some drugs through pharmaceutical company advertising, sales representatives, or other marketing strategies [45,52].
- Lack of promotion of highly effective but nonprofitable medications (e.g., spironolactone for heart failure) [45,52].
- Simple errors of omission [8,23,25,48,52].
- Negative attitudes toward issues of cost effectiveness of medications.
- Direct-to-consumer marketing strategies and other competing influences [49].
- Patient and family demand for a particular agent, even when it is not scientifically substantiated [49,50,52].
- Physician overreliance on clinical experience in opposition to scientific data [50,51].
- Skepticism toward, and distrust of, the literature and academia among some community-based physicians [51].
- Clinical inertia [52].
- The need to take some definitive therapeutic action even when "watchful waiting" may be the most justifiable action [50,52].
- Concerns related to medicolegal liability and the perceived need to practice defensive medicine [37,50,51].
- Influence from clinical opinion leaders or other health practitioners [50–52].

These diverse influences suggest the need for tailoring multifaceted intervention strategies to the key factors influencing a given clinical behavior, based on models of behavioral change and knowledge translation.

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

Research on the impact of educational and administrative interventions to improve drug prescribing presents numerous methodologic challenges. This section will review several of the most important methodologic problems,

such as internal validity, regression toward the mean, unit of analysis errors, logistical issues, ethical and legal problems, and the detection of effects on patient outcomes.

Internal Validity

As early as 1975, Gilbert *et al*. established that poorly controlled studies produce misleading estimates of the effects of a variety of social programs [53]. Many nonintervention factors can affect medication use over time, such as marketing campaigns, mass media, state or federal regulatory policies, seasonal effects, changing staff of healthcare organizations, other "competing" interventions, changes in eligibility for insurance programs, shifting demographics, and so on. Because RCTs are sometimes not feasible (e.g., contamination of controls within a single institution) or ethical (e.g., withholding quality assurance programs from controls), other strong quasi-experimental designs (e.g., interrupted time series with or without comparison series, pre–post with concurrent comparison group studies) should be used instead of weak one-group post-only or pre–post designs that do not generally permit causal inferences. In fact, the Cochrane Collaboration's EPOC Group considers rigorously conducted time-series studies and pre–post studies with a concurrent comparison group (and several baseline observations to control for secular trends) to be sufficiently valid to merit inclusion within its systematic reviews [54].

Interrupted time-series designs include multiple observations (often eight or more) of study populations before and after intervention. Such designs often permit investigators to control for preintervention secular changes in study outcomes, and to estimate the size and statistical significance of sudden changes in the level or slope of the time series occurring at initiation of the treatment. The availability of a comparison series collected from a similar, but unexposed, comparison group can further increase causal interpretability if no simultaneous change in trend is observed for this group [18,55].

Another popular design that may lead to valid results is the *pre–post with comparison group design*. Ideally, this design includes several observations before and at least one observation after treatment in a nonrandomly selected group exposed to a treatment (e.g., physicians receiving feedback on specific prescribing practices) [54], as well as simultaneous before-and-after observations of a similar (comparison) group not receiving treatment. Although this design controls for many threats to the validity of causal inferences (e.g., due to the effects of testing or maturation), it sometimes cannot control for unknown factors (e.g., a regulatory policy), which might result in preintervention differences in trends between study and comparison groups, and it is thus not as rigorous as a controlled interrupted time-series design [53,55].

The weakest, and not uncommon, design is the *one-group, post-only design*, which consists of making only one observation on a single group that has already been exposed to a treatment. The *one-group pre–post design* merely adds a single preintervention observation to the previous design. Such weak designs are unlikely to produce valid or reliable estimates of the effects of interventions, so much so that they are routinely excluded from careful reviews of the literature [18–24,54]. Furthermore, many (if not most) studies of newer technology-based approaches to improving prescribing, such as computerized physician order entry and other types of computerized decision support, have used the post-only or one-group pre–post designs to evaluate their efficacy and effectiveness [56–58].

Inadequately controlled studies may exaggerate the effectiveness of many interventions to improve prescribing. For example, inadequately controlled studies of the dissemination of print-only materials used alone have all reported positive effects on behavior, while well-controlled

studies of such strategies all reported small or nonexistent changes in behavior [59]. The "success" of uncontrolled studies is often due to the attribution of preexisting trends in practice patterns to the studied intervention.

There are many examples of the potential bias involved in failing to account for prior trends. In one study, the naturally occurring trends in the use of 23 categories of medication were examined in a four-year study of 390 000 enrollees in the New Jersey Medicaid program [60]. The results indicated that 50% of the estimated one-year percentage changes in prescriptions per 1000 enrollees exceeded +20.3% or −10.8% of baseline levels. Effect sizes reported in the prescribing intervention literature are similar to these natural fluctuations [61], suggesting that changes in drug use attributed to such interventions could merely reflect these underlying secular trends. This is particularly noteworthy, because the effect sizes reported for valid intervention studies tend to be modest at best, with improvements in the quality of prescribing (as variously defined by investigators) usually reported on the order of a 10–20% absolute improvement over controls.

Recently the Institute for Health Improvement (IHI) in the US failed to control for baseline trends before its nationwide hospital safety program, the "100,000 Lives Program," and falsely stated that the program saved 130 000 lives until a reanalysis including a long baseline showed already occurring mortality changes before the policy (Figure 19.1) [62].

These examples provide further support for more widespread application of RCTs or, when RCTs are not feasible, time-series and other valid comparison series designs to evaluate

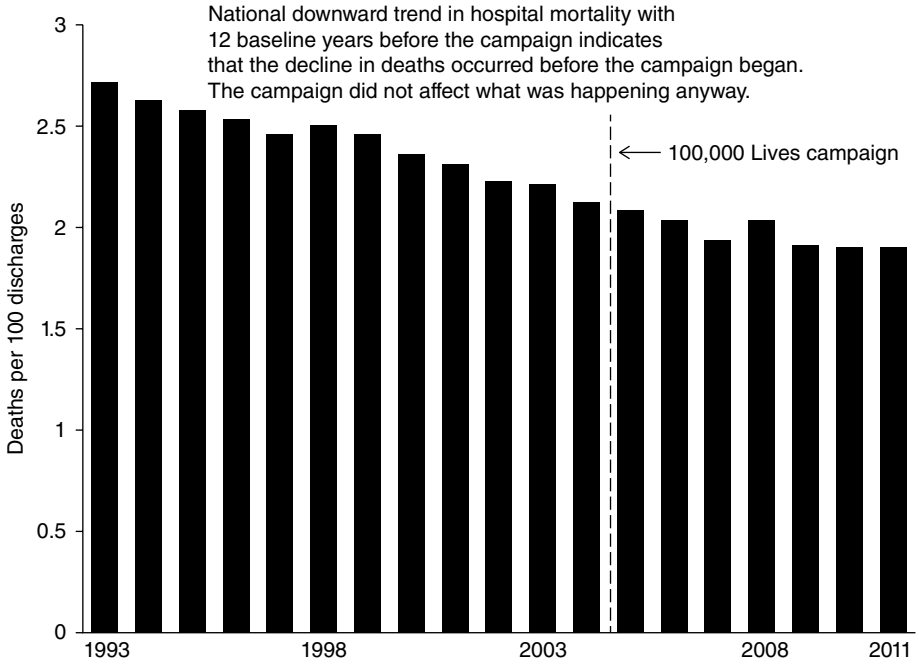


Figure 19.1 Example of a strong time-series design that controlled for history bias in the Institute for Healthcare Improvement's 100,000 Lives Campaign. Figure is based on data from the Agency for Healthcare Research and Quality.

whether suddenly introduced interventions are associated with corresponding changes in the level or slope of the utilization series, after controlling for prior trends. If the collection of time-series data is not feasible, investigators may consider using pre–post with comparison group designs (ideally incorporating several baseline points), which also control to some degree for temporal changes and unforeseen co-interventions that may concurrently affect prescribing or utilization, as described in respected texts on intervention research design [53,55].

Regression toward the Mean

Regression toward the mean – the tendency for observations on populations selected on the basis of exceeding a predetermined threshold level to approach the mean on subsequent observations – is a common and insidious problem in much of the drug utilization literature. For example, the most common Medicaid DUR programs typically screen prescribing data and eligibility files for possible co-occurrences of two interacting medications, or higher than recommended dosages for individual drugs. After case-by-case review by expert committees, letters (or email equivalents) are written to responsible physicians questioning the practice and asking for written responses. However, methodologic papers on this topic indicate that published research evaluating DUR used poorly controlled designs that were unable to control for regression to the mean [14,17,18]. For example, in one often-cited DUR study [61], 50% of prescribing problems were absent several months after letters were sent, suggesting to the noncritical reader that the program was effective. However, it is equally plausible that the offending medications were withdrawn because the patients' conditions improved or because the physicians detected the error on their own.

The likelihood that all screening algorithms employed in DUR programs are subject to regression toward the mean argues strongly for the need to conduct RCTs and well-controlled quasi-experiments (e.g., pre–post with comparison group design) to justify the efficiency and effectiveness of these interventions *before* they become a routine part of private and public quality improvement programs [14,17,18]. If regression effects are unavoidable – for example, due to selection of at-risk populations – investigators may consider including a "wash-out" period after selection and before pre- and postintervention observations [18,46].

Unit of Analysis

A common methodologic problem in studies of physician behavior is the incorrect use of the patient as the unit of analysis [63–66]. Such a practice violates basic statistical assumptions of independence, because prescribing behaviors and outcomes for individual patients are likely to be correlated within each physician's practice. To some degree, the prescribing practices of physicians within a group practice may also not be statistically independent of each other [64–66]. These forms of hierarchical "nesting" or statistical "clustering" often lead to accurate point estimates of effect but inappropriately low P values and narrow confidence intervals, when the unit of analysis is assumed to be a statistically independent patient and the analytic framework does not account for correlation among patients treated by the same physician, or groups of physicians within a practice or hospital [64–66]. As a result, interventions may appear to lead to statistically significant improvements in prescribing practices because of mistakenly inflated sample sizes. For example, one review of articles on physicians' patient care behavior found that 70% of 54 articles incorrectly analyzed the data using the patient as the unit of analysis without accounting for statistical clustering; among 19 reviewed studies of medication prescribing, 58% used the incorrect unit of analysis [66].

The simplest, although overly conservative, solution to the problem of incorrect unit of analysis is to analyze data by facility or physician. Fortunately, more statistically efficient methods for analyzing clustered data are becoming increasingly available; such models can simultaneously account for clustering of observations at the patient, physician, and facility levels [65–70]. Such models allow aggregation at the patient level by accounting for correlation between patients cared for by the same provider or facility. The resulting P values for differences in prescribing rates between study and control groups are almost always more conservative (and confidence intervals wider) than assuming no intraclass correlation, but are still greater (narrower confidence intervals) than the most conservative methods of analyzing at the provider or facility level. Much methodologic work remains to be done in terms of understanding what the appropriate unit of allocation and analysis is for various studies, how to best estimate power and sample sizes, and whether sensitivity analyses regarding unit of analysis need to be conducted or presented in the results of such studies.

Logistical Issues

While continuity of care is a goal in most healthcare settings, many patients, particularly those treated within academic medical centers, see multiple primary providers over time. For example, patients treated by residents may be reassigned to other residents at the end of the academic year. Providers may go on extended leave and transfer cases to other clinicians. Patients themselves may choose another primary care provider. In addition, many patients develop ongoing relationships with specialists as particular problems develop and are resolved.

While these changes may or may not affect patients' care, they almost always complicate and sometimes weaken research conducted in a clinical setting. Particularly in settings where providers may be assigned to both "intervention" and "control" patients, contamination problems are difficult to avoid. Even when interventions can be focused effectively on the intended patients or providers, informal communication among providers can lead to contaminated effects, thereby decreasing the likelihood of detecting significant changes.

Fortunately, some solutions to these problems exist. First, investigators should identify, through baseline interviews and organizational records, the extent to which patients are cared for by multiple providers, and the patterns of consultations and referrals between caregivers within and between facilities. If randomization of clinicians is likely to lead to contamination of controls, or if patient–provider pairs are frequently broken, then randomization of facilities should be used, such that an entire facility or subunit cluster (e.g., the "firm" within an academic teaching hospital or the "primary care practice" in the community) is assigned to the same study group. For instance, a quality improvement intervention cluster randomized 37 hospitals in one state to intervention or control status [69]. However, when this strategy is not feasible, because it results in a small sample of facilities and inadequate statistical power, investigators are encouraged to collect data during multiple observation periods both before and after the intervention, and to use time-series regression methods that can often detect modest changes in utilization levels after as few as 6–12 months.

Ethical and Legal Problems Hindering the Implementation of Randomized Clinical Trials

Adequate control groups are essential for rigorous evaluation of results. Yet it has been argued that there are ethical and legal problems related to withholding interventions designed to improve drug prescribing practices. This argument explicitly assumes that the proposed

interventions are known to be beneficial. In fact, the efficacy and effectiveness of many programs to improve drug use is the very question that should be under investigation. Some commentators have argued, quite reasonably, that mandating such programs or interventions without adequate and valid proof of benefit is in fact unethical. For example, many researchers and policymakers have stated that computerized physician order entry (CPOE) does not need to be studied, and the Leapfrog advocacy group has gone so far as to state that not having CPOE compromises patient safety and quality of care [71]. What is important is to demonstrate that such interventions are safe, efficacious, and cost-effective *before* widespread adoption. Later studies have, in fact, found that health information technology can cause unintended patient harms or deaths [72,73]. Even a safe and nonefficacious intervention is associated with opportunity costs and unknown harms; if this given intervention is widely adopted or legislatively mandated, many resources will have been diverted away from other parts of the healthcare delivery system. In those very rare instances in which the intervention has shown unusual promise in similar populations, the application of RCTs may be inappropriate, but alternative research designs should still be considered to better define the absolute risks, benefits, and costs of the intervention. Feasible design alternatives are quasi-experimental designs such as interrupted time-series analysis or staged implementation, in which the control population (or regions) receives the intervention after comparative data have been collected [29,55,65,70,74,75].

Detecting Effects on Patient Outcomes

While a number of studies have demonstrated beneficial effects of various interventions on prescribing practices, few large well-controlled studies have linked such changes in the processes of care to improved patient outcomes. A notable exception was a randomized trial of computerized alerts to improve venous thromboprophylaxis for hospitalized patients in a teaching hospital of over 700 beds [75]. Kucher *et al*. allocated about 2500 patients and their physicians to either usual care or exposure to a computerized alerting system that automatically generated a clinical risk of deep venous thrombosis score, and alerted physicians to the need for prophylaxis using either drugs or devices. Unlike most studies, this trial was designed to detect a difference in clinical events, namely objectively diagnosed life-threatening deep venous thrombosis or pulmonary embolism [75]. The computerized system more than doubled rates of appropriate prophylaxis (34% vs. 14% for controls), although there was still room for improvement. More importantly, it led to a clinically important (41% decrease) and statistically significant ($P = 0.001$) reduction in adverse clinical events. There are only a limited number of studies suggesting a tight link between improvements in processes of care and patient-related outcomes. Under most circumstances, it is profoundly difficult to demonstrate statistically significant changes in patient outcomes in response to intervention. Explanations for the far more commonly observed dissociation between improvements in prescribing and better patient outcomes include that (i) easily available clinical measures (e.g., mortality, unplanned hospital admission) may not be sensitive to the kinds of patient-related outcomes that might be affected by introduction or withdrawal of medications; (ii) changes in physician prescribing may lead to little or no change in patients' health status if patients do not adhere to the recommended regimens; and (iii) many medical therapies require months to years of continued persistence before clinical benefits become apparent.

Because of these problems, sample sizes may need to be enormous to detect even very modest changes in patient outcomes (see Chapter 4

for a discussion of methods for determining statistical power). These problems are much less severe in drug trials (especially placebo-controlled studies) because of experimenter control over the major independent variable – exposure to medications (see Chapter 36). However, process outcomes (e.g., use of recommended medications for acute myocardial infarction from evidence-based practice guidelines) are often sensitive, clinically reasonable, and appropriate measures of the quality of care [74–76], and improvements in process should not be dismissed outright as surrogate outcomes. They may be important in and of themselves, as long as the processes are a measure of evidence-based and proven effective therapy [74–76].

Conceptual Framework for Changing Clinical Behaviors

A useful starting point for designing an intervention to improve prescribing is to develop a framework for organizing the clinical and nonclinical factors that could help or impede desired changes in clinical behaviors [7–10,77]. The theory of planned behavior [9,10] is amenable to developing such a framework, as is the predisposing, enabling, and reinforcing (PRECEDE) model [77]. PRECEDE was developed for adult health education programs by Green and Kreuter [77], and proposes factors influencing three sequential stages of behavior change (predisposing, enabling, and reinforcing factors). *Predisposing* variables include such factors as awareness of a consensus guideline on appropriate use of a thrombolytic agent, knowledge of clinical relationships supporting such a guideline (e.g., major actions of thrombolytics in the artery), beliefs in the efficacy of treatment (e.g., probability of survival), attitudes or values associated with recommended behaviors (e.g., risk of intracranial hemorrhage associated with therapy), and a myriad of other potential factors [8,52]. However, while a mailed drug bulletin or email alert may predispose some physicians to new information (if they read it), behavior change may be impossible without new *enabling* skills (e.g., skills in administering a new therapy, or overcoming patient or family demand for unsubstantiated treatments). Once a new pattern of behavior is tried, multiple and positive *reinforcements* (e.g., through peers, reminders, feedback, and incentives) may be necessary to establish the new behavior fully. Several reviews of the literature have come to a similar conclusion [20,34,52,78]: multifaceted interventions that encompass all stages of behavior change are most likely to improve physician prescribing.

Empirical Evidence on the Effectiveness of Interventions to Improve Prescribing

Does existing empirical evidence on the effectiveness of alternative prescribing interventions provide any lessons on the key characteristics of successful approaches to this problem? Illustrative findings from several research syntheses will be used to evaluate the effectiveness of the most commonly studied or applied approaches. Because of severe biases introduced by uncontrolled designs that do not measure preexisting trends in target drug use behaviors (see prior "Methodologic Problems" section), only studies using valid experimental or quasi-experimental research designs (e.g., RCTs and time-series designs) are discussed.

Educational Interventions

Printed Educational Materials and Guidelines

Distributing printed educational materials aimed at improving prescribing practice remains the most ubiquitous form of prescribing

education in the industrialized world. While the most sophisticated materials may incorporate visually arresting graphs, illustrations, and headlines to convey important behavioral and educational messages, such a strategy rests on assumptions that physicians will be exposed to the information, and that such rational information will be sufficiently persuasive to change clinical practices. Unfortunately, several reviews provide consistent evidence that use of disseminated educational materials *alone* (such as drug bulletins, self-education curricula, objective, graphically illustrated "un-advertisements," or other professionally prepared educational brochures) may affect some of the predisposing variables in the change process, but will have a minimal (and often no) effect on actual prescribing practice [11,19,23,52,78–81].

Clinical practice guidelines are a distinct subset of educational materials. Although primarily educational in nature, they are also a codification of current best practice, and are intended to improve quality and decrease costs by minimizing unnecessary variations in practice. However, faith in the simple act of guideline dissemination presupposes that information alone, regardless of how reliable or well referenced, can change behavior. In general, when rigorously studied, guideline dissemination *alone* does not influence prescribing behavior or other practices to a clinically important degree [11,19,34,80–86]. Given the proliferation and availability of numerous guidelines, dissemination of a particular guideline should be considered part of "usual care," and so unlikely to change practice, and to provide a reasonable control "intervention" with which to compare more effective interventions or strategies. The failed IHI 100,000 Lives Program cited earlier also relied only on printed guidelines to avert hospital deaths [87].

In summary, simple dissemination of educational materials does not appear to be effective by itself in altering prescribing patterns, but these materials may provide a necessary *predisposing* foundation for other *enabling* and *reinforcing* strategies.

One-to-One Education (Academic Detailing)

A number of controlled studies support the conclusion that programs with brief face-to-face visits (15–25 minutes) by pharmacists, physician counsellors, or peer leaders (academic detailers) are effective in promoting evidence-based practice and improving patient outcomes [11,19,88–91]. For instance, a Cochrane review of 30 trials suggests a 6% increase in compliance with practice guidelines when any intervention in which educational meetings were a component was compared to no intervention [19], and another indicates interventions involving peer leaders were associated with a 12% increase in compliance with practice guidelines [88]. The principles and methods of this approach are described in detail elsewhere [12], and include (i) targeting of physicians with higher than average needs for education (e.g., through analyses of administrative data); (ii) conducting motivational research (e.g., surveys of focus group interviews) in advance of the intervention to understand the causes of suboptimal prescribing patterns; (iii) sponsorship by authoritative and credible medical organizations; (iv) two-way communication with prescribers to increase clinician involvement and relevance to different patient populations and settings; (v) presentation and discussion of counterarguments to which physicians have been exposed; (vi) brevity; (vii) use of high-quality, graphical educational materials; (viii) repetition of major messages; and (ix) follow-up visits for positive reinforcement. Of course, pharmaceutical industry detailing also shares many of these principles and methods. What sets academic detailing apart from industry efforts is that the messengers and the messages of the former are independent, objective, and evidence based.

A formal economic analysis of academic detailing in this study, conducted from a societal

perspective (in this case, a Medicaid program), concluded that targeting moderate to high prescribers of propoxyphene, cephalexin, and vasodilators using administrative claims databases could lead to a benefit-to-cost ratio of more than 2, even without considering positive spillover effects to nonparticipating physicians, improved quality of care, or possible cost savings due to elimination of adverse drug effects [84]. The main barrier to more widespread use of the strategy is its *perceived* labor intensiveness. Nevertheless, academic detailing is a consistently effective method for changing physician practice [11,19,88,90–93].

Group Education and Group Detailing

Although rounds, seminars, and other group didactic educational programs are among the most universal methods for prescribing education, controlled studies of this approach are almost nonexistent in the literature, especially in nonteaching settings. Nevertheless, small group discussions conducted by clinical leaders in academic primary care settings have been shown to improve use of antibiotics [40] and agents for hypertension treatment and control [94]. These successful approaches have included reviews of patient records to establish the need for change and participatory methods based on adult learning theory, and have more in common with academic (individual or group) detailing than traditional modalities of continuing medical education. Traditional large-group, didactic continuing medical education seminars have not been as successful, by themselves, in improving physician performance [19,20,95,96]. Even the most rigorous internet-based extensions of traditional continuing medical education had yielded only negligible incremental advantage over more traditional approaches [96]. The results of one early but important RCT of continuing medical education were summed up by the authors as follows: "Put simply, in terms of the effects of continuing education on the documented quality of care, wanting continuing education … was as good as getting it" [95].

A number of controlled trials have replicated the positive results of one-to-one outreach with smaller group outreach sessions, often referred to as "group detailing" [97]. Group detailing has the potential advantage of encouraging discussions within the group, which may enhance the diffusion of ideas and increase their impact. For example, in a cluster RCT to improve the use of antihypertensive medications in primary care in the US, Simon *et al.* randomly allocated three practice sites to group detailing (n = 227 prescribers), three to individual detailing (n = 235 prescribers), and three to usual care (n = 319 prescribers). Individual detailing entailed a physician-educator meeting individually with clinicians to address barriers to prescribing guideline-recommended medications. The group detailing intervention incorporated the same social marketing principles in small groups of clinicians. Results of this study suggested that both group and individual academic detailing improved antihypertensive prescribing (adjusted odds ratio [OR], 1.40; 95% confidence interval [CI], 1.11–1.76 and adjusted OR, 1.30; 95% CI, 0.95–1.79, respectively) compared to usual care, but may require reinforcement to sustain improvement [98].

It is likely that as long as the group size is kept relatively small (i.e., fewer than 5–10 participants), and the other precepts of academic detailing are adhered to, group detailing is a reasonable alternative approach to individualized educational outreach.

Local Opinion Leaders

The role of local opinion leaders in the adoption of new pharmaceutical agents has been well documented by Coleman *et al.* [2]. Their data indicated that after opinion leaders adopted particular drugs, other less integrated physicians eventually followed in a classic curve of technology diffusion. In several studies of the diffusion of scientific information on treatment

of arthritis and the inappropriate use of Cesarean sections [85,99], local opinion leaders or educationally influential physicians have been identified and encouraged to consult informally with colleagues. These opinion leaders are approached frequently for clinical advice, are trusted by their colleagues to evaluate new medical practices in the context of local norms, have good listening skills, and are perceived as clinically competent and caring [3,69,85,89,99–101]. In addition to opinion leader involvement, these interventions generally included a brief orientation to research findings, printed educational materials, and encouragement to implement guidelines during informal "teachable moments" that occur naturally in their ongoing collegial associations. The success of these programs was attributed to "the importance of the local community's norms, the orientation of practitioners to locally credible individuals, and the need to translate the research findings into a locally applicable message" [100]

Although the recruitment and use of opinion leaders show great promise in accelerating the adoption of evidence into practice, overall the results of rigorous opinion leader studies have been mixed [88], and whether or not such interventions are reproducible across diseases and settings [101], can improve prescribing for multiple conditions outside the hospital setting, and are cost effective remains to be determined.

Monitoring and Feedback

Prescribing, Audit, and Feedback

A popular approach to improving physician performance has been some form of "feedback" of prescribing patterns to individuals or groups of physicians. It has been estimated that, annually, more than one-half of all US physicians receive some clinical or economic feedback regarding their prescribing practices [25,102].

Audit and feedback interventions often compare practice patterns with peers or predetermined standards such as practice guidelines. The former is typified by interventions of peer comparison feedback, while the latter are typified by formal drug utilization review programs. Systematic reviews consistently conclude that peer comparison feedback has a statistically significant, but clinically minimal, effect on prescribing or other physician behaviors [11,21,25,102]. For instance, a Cochrane review of 49 trials suggests a 4.3% increase in healthcare professionals' compliance with practice guidelines and a 0.4% decline in dichotomous patient outcomes (e.g., the proportion of patients with appropriate management) [102].

Furthermore, it seems these programs would be unlikely to offset the costs of the interventions themselves, much less lead to cost savings. The conclusions of these reviews were supported by a methodologically rigorous RCT of the effect of peer comparison feedback on the prescription of five unrelated groups of medication [103]. These Australian investigators went so far as to conclude, based on their null results, that "feedback is not worthwhile and should not be seen as a high priority by government agencies" [103].

In addition to the type or content of the feedback, a number of variables must be considered. Communication channels could be by letter, computer, or face-to-face encounter with a supervisor or colleague. Even more importantly, the credibility of the source of the feedback information probably influences its effectiveness more than the content of messages. Thus, feedback programs operated by a government regulator or managed care organization may be less effective than professionally based educational programs in which an ongoing relationship exists between the sender and receiver of information [103–108].

Lastly, if physicians are not able to respond immediately to the feedback delivered, by altering prescribing during a specific patient encounter, they may not respond at all. It is not

necessarily true that physicians will generalize behavior from one specific encounter to similar clinical situations [104,105].

One important advance in the area of audit and feedback, which attempts to address many of the aforementioned problems, is the development of the concept of the "achievable benchmarks of care" by Kiefe *et al.* [105]. The underlying theory is that viewing one's personal performance within the context of peers' performance should be a powerful motivator for change [89]. In essence, the achievable benchmark represents the average performance of the top 10% of local physicians being assessed [105]. By design, achievable benchmarks are higher than the group mean – and group mean data are what are most often provided in audit and feedback programs. Kiefe *et al.* undertook a cluster RCT and allocated physicians (n ~ 100) and their diabetic patients (n ~ 2000) to either "usual care" (in fact, it was a standard quality improvement intervention that profiled physicians and provided them with individual and group mean performance feedback on five different quality indicators, such as influenza vaccination, foot examination, and measurement of glycosylated hemoglobin) or to an experimental intervention (usual care plus the provision of top 10% achievable benchmark data). The intervention was associated with 15–57% relative improvements in all indicators compared with usual care; three out of five of these improvements were also statistically significant [105].

Drug Utilization Review

DUR programs have been defined as "structured, ongoing initiatives that interpret patterns of drug use in relation to predetermined criteria, and attempt to prevent or minimize inappropriate prescribing" [14,17]. DUR has many synonyms, including drug use review, drug use evaluation, and medication review. It involves a comprehensive review of patients' prescription and medication data before, during, and after dispensing to ensure appropriate medication decision making and positive patient outcomes. DUR programs provide prescriber feedback at the patient level.

Prospective or retrospective DUR are well-studied forms of prescriber feedback. It has frequently been hypothesized that simply making clinicians aware of all of the medications a patient may be prescribed might be an effective method for reducing use of excessive, duplicative, or interacting medications. The best controlled trials of this approach confirm that simply distributing such profiles, without explicit suggestions for changes in practices, has no detectable effect on prescribing practice [103,109,110]. Likely reasons for the failure of this intuitively appealing approach include that (i) much of the generated information was probably clinically irrelevant; (ii) unsynthesized and voluminous data may cause information "overload" and desensitization of busy clinicians; (iii) there was no provision of alternative measures to improve care; and (iv) the feedback was not derived from credible sources of information. This approach represents one of the few instances in which the volume of negative findings from methodologically rigorous studies strongly supports the exclusion of this strategy from future research.

Many existing DUR programs attempt to review the appropriateness of medication prescribing for individual patients (e.g., drug interactions and dosage). Since the majority of feedback messages are likely to be clinically unimportant [14,17,18,103], the clinically relevant messages could be unintentionally ignored. Recent systematic reviews of medication review in adult hospitalized patients and nursing home residents found no evidence that medication review reduces mortality or hospitalizations [111,112]. In fact, the best-controlled studies do not yet support the effectiveness of either retrospective or prospective DUR, even though both are mandated for all state Medicaid programs [14,17,18].

Reminders and Computerized Decision Support Systems

Often, physicians are predisposed to certain therapeutic interventions, but simply omit them due to oversight or lack of coordination in the healthcare/communications system. In these cases, computerized reminders can enable physicians to reduce these errors of omission by issuing alerts to perform specific actions in response to patient-level information such as laboratory findings or diagnoses.

Several studies in hospitals, managed care organizations, and primary care settings have provided strong evidence that such systems can prevent the omission of essential preventive services such as deep venous thrombosis prophylaxis, influenza immunization, and others [57,75,113–115]. It is noteworthy that computerized reminders and decision support systems (see later discussion) are perhaps the most-studied aspect of the various functionalities present within the electronic health record. A rigorous review of 28 trials examining real-time point-of-care computerized reminders demonstrated that, when compared with controls, the median improvement in process-of-care measures was only 4%; the median improvement in appropriate medication prescribing was only 3% (interquartile range 1–11%) [113]. Although statistically significant, the magnitude of improvements is far smaller than expected and in many cases might not be considered worthwhile. Other systematic reviews [114,115] also concluded that reminders probably slightly improve quality of care, in terms of compliance with clinical guidelines, but little evidence exists on whether reminders improve patient outcomes. The effects of reminders on improving quality of care vary across settings and conditions. Such systems are more likely to be effective if the reminder requires a response from the clinician and provides an explanation of the reminder's content or advice [114]. "Reminder fatigue," with concurrent bypassing of computer screens or generalized neglect of all alerts, is an important problem that has not been well addressed [116,117].

A major component of health information technology (HIT) is computerized decision support systems (CDSS) integrated with electronic health records. They are intended to support physicians' prescribing decisions at the point of ordering, including alerts regarding dosage, drug interactions, schedule, suboptimal choices, and prevention of adverse drug events [113,118,119]. This promise, however, should not be assumed [57,72,73,113,119]. In one older but very rigorous study of advanced computer decision support, Eccles *et al*. conducted a cluster RCT of 60 busy primary care practices in the UK [10]. These practices already had electronic records and electronic prescribing. Eccles *et al*. randomized them to a computerized guideline/decision support intervention that was fully integrated into the electronic clinic record; half of the practices were allocated to a symptomatic coronary disease guideline (n = 1415 intervention patients) and the other practices to an asthma guideline (n = 1200 intervention patients). After one year, there were no significant improvements in any one of more than 40 different quality indicators for either condition [10]. Available systematic reviews suggest that CDSS might moderately improve process of care such as rates of laboratory monitoring and prescribing decisions, and that they may help reduce the length of hospital stay compared with routine care while comparable or better cost-effectiveness is achieved, but there is no evidence that CDSS have fulfilled their promised effect on healthcare costs, mortality, or other clinical adverse events [120–124]. Furthermore, there have also been well-documented harms and adverse events induced by various CDSS [72,73]. With these cautionary notes, we refer the interested reader to a more detailed examination of adverse drug events in general, and the potential roles of CPOE and computerized decision support, in recent systematic reviews [57,108,113,120–123,125].

Multimedia Campaigns

Occasionally, the discovery of important adverse effects of marketed drugs is accompanied by dissemination of educational materials to physicians as part of a broader warning campaign involving the medical and popular press, internet, newspapers, television, and radio. When the adverse effects are severe and preventable, alternative agents exist, and the messages are simple enough to convey in mass communications, such multimedia campaigns may be effective in changing prescribing patterns in large populations.

Figure 19.2 provides data from a US study suggesting that widespread reporting by the medical and lay press of the risk of Reye's syndrome associated with pediatric aspirin use was associated with declines in Reye's syndrome incidence. This media campaign was conducted after several epidemiologic studies identified the association between Reye's syndrome and aspirin use and antecedent viral illnesses [126]. The authors concluded, based on this and other studies, that mass media warnings may be effective in changing both consumer and physician behavior when the illness is severe or life threatening, the behavioral message is simple, no or few barriers to alternative behaviors (e.g., acetaminophen versus aspirin) are present, and the campaign is comprehensive, involving both health professionals and consumers.

Figure 19.3 provides data from another example of the effects of multimedia campaigns. A recent study [127] examined the effects of widespread reporting by the medical and lay press of suicidality risk associated with pediatric antidepressant use. In a 10-year interrupted time-series analysis in 11 health plans across the US, the authors found that the drug safety warnings and hyped media coverage led to substantial reductions in antidepressant use and small, visible increases in psychotropic drug poisonings treated in emergency rooms and hospitals.

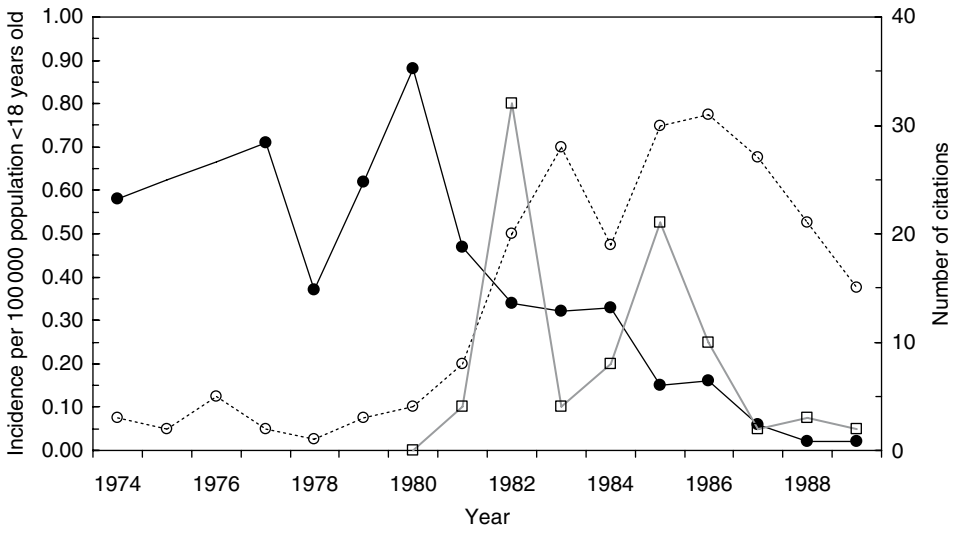


Figure 19.2 Trend in number of (○) medical and (□) lay press citations on aspirin and Reye's syndrome, and the incidence of Reye's syndrome (●) among children. Newspaper index limited to four continuously reporting national newspapers described in text. *Source:* Soumerai SB, Ross-Degnan D, Spira J. The effects of professional and media warnings about the association between aspirin use in children and Reye's syndrome. *Milbank Q* 1992; **70**: 155–82. Reprinted by permission of *Milbank Quarterly*.

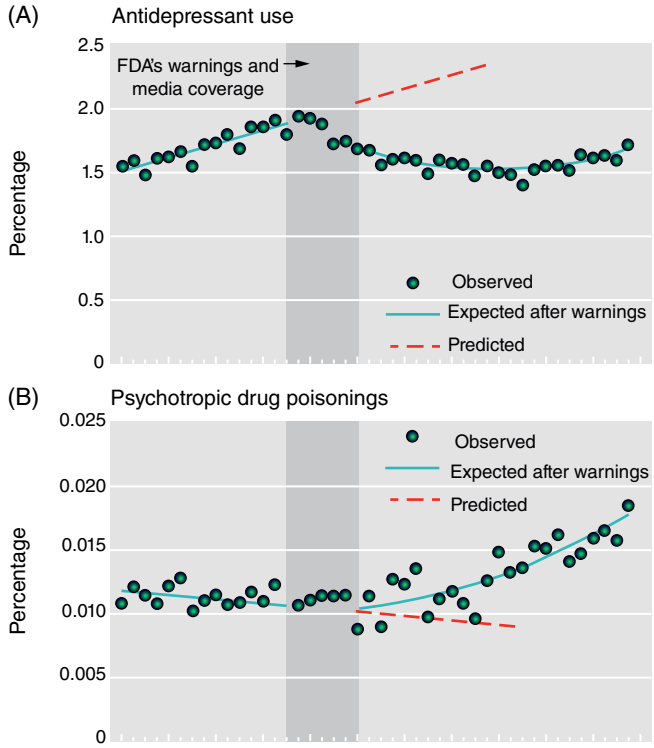


Figure 19.3 Rates of (A) antidepressant use and (B) psychotropic drug poisonings per quarter before and after the warnings among adolescents enrolled in 11 health plans in a nationwide Mental Health Research Network. *Source:* Lu CY, Zhang F, Lakoma MD, *et al*. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: quasi-experimental study. *BMJ* 2014; **348**: g3596. Reproduced with permission of BMJ Publishing Group Ltd.

Formulary Interventions and Financial Incentives

In response to escalating drug costs, an increasing number of physician practices in the US are entering into capitated drug risk-sharing arrangements with managed care organizations [22–24]. These innovative drug payment mechanisms are designed to control drug costs by encouraging physicians to prescribe "preferred" drug products (e.g., generic drugs or those that are on the health plan's formulary). Some analysts assert that capitation encourages physicians to examine their prescribing more critically, resulting in the choice of appropriate, effective, and low-cost medications. This belief is based on a number of untested assumptions: (i) practices are large enough to absorb risk, so that costly but appropriate prescribing decisions for the individual patient are not unduly affected; (ii) performance feedback to prescribers is timely and provides specific advice about costs, risks, and possible substitutions; and (iii) physicians understand and are sensitive to differences in drug pricing. Because it is unlikely that these assumptions (in general) are met, any intervention using financial incentives must be considered experimental. This is perhaps most true for various "pay-for-performance" schemes that have been introduced in many settings [128–130]. The most mature pay-for-performance program is that linked to the UK Quality and Outcomes Framework for primary care [128]. There were more than 100 indicators introduced, and by most standards the incentives were considered generous – up to an additional 25% of an individual physician's income could be generated within the scheme, and collectively the equivalent of about $1 billion annually

was spent by the National Health Service. Rigorous studies, using time-series methods, demonstrated small, if any improvements in performance indicators; protracted follow-up, however, demonstrated that performance plateaued and in fact quality of care may have even decreased for those conditions or indicators that were not incentivized [128–130]. In a recent RCT [131] of a pay-for-performance intervention, physicians were randomized to receive economic incentives for each patient who met target cholesterol levels, whereas physicians in the control groups received no economic incentives for achieving better outcomes. Results of this RCT suggest that physician payments did not produce any meaningful changes in quality of care compared with an equivalent group receiving no incentives.

Preferred drug lists (PDLs) are commonly used in the US and elsewhere by policymakers to contain drug costs; prescribing of nonpreferred drugs requires prior authorization. Well-controlled studies have assessed their effects on the quality of prescribing, unintended consequences, and patient outcomes. Rigorous, longitudinal studies reported that prior authorization is associated with lower use of nonpreferred cardiovascular drugs [132,133], and with treatment discontinuity among patients with severe mental illness such as schizophrenia and bipolar illness [134–136] with few cost savings. Following treatment discontinuation associated with prior authorization, a pre–post design with a control group study [137] found significant and concerning reductions in psychiatric visits among the sickest patients with bipolar illness and increases in emergency room visits among less severely ill patients, which may reflect attempts by patients to improve access to medication. An interrupted time-series study found that prior authorization was associated with substantial reductions in initiation of nonpreferred medications [138] for severe mental illness without offsetting increases in use of preferred agents. These results suggest that prior authorization acts as an administrative barrier and is associated with underuse of maintenance medications. This is a major concern for the quality of care of this population, because underuse of maintenance medications could lead to acute episodes of illness, suicide, and hospitalization. Moreover, targeting essential drug classes with heterogeneous patient responses and side effects could reduce appropriate care, adversely affect health status, and cause shifts to more costly types of care. Assessing inappropriate use of high-cost drugs before implementing regulations and instituting simple mechanisms to exempt high-risk patients could maximize savings and minimize harm.

The Future

Based on this synthesis of the research literature, it is clear that our knowledge of the characteristics of successful interventions to improve prescribing is growing rapidly. Passive dissemination of evidence is a necessary but insufficient method for improving most prescribing behaviors. In general, the achievement of long-term changes in practice will depend on the inclusion of multiple strategies that predispose, enable, and reinforce the desired prescribing behaviors. The following characteristics seem to recur in successful interventions:

- Using theoretical and conceptual frameworks to identify key factors influencing prescribing decisions through surveys, focus groups, or in-depth interviews.
- Targeting physicians in need of education (e.g., through review of prescribing data) to increase effectiveness and efficiency.
- Recruitment and participation of local opinion leaders.
- Use of credible and objective messengers and materials.

- Face-to-face interaction, especially in primary care settings.
- Audit and feedback (*if they are used at all*) that incorporate achievable benchmarks, comparisons with peers, and patient-specific data.
- Repetition and reinforcement of a limited number of messages at one time.
- Provision of acceptable alternatives to the practices that are deemed necessary to be extinguished.
- Brief, graphic educational guidelines and evidence summaries to predispose and reinforce messages.
- Use of multiple evidence-based strategies to address multiple barriers to best practice.
- Emphasis on the goal of improvement in the quality of prescribing and patient safety, not just cost minimization in the guise of quality improvement.

There is also a tremendous need for carefully controlled research of some existing and new methods for improving prescribing, and how best to combine various evidence-based strategies to allow for rapid local implementation of prescribing guidelines. New models are needed to predict the most effective types of intervention for specific problem types, and a number of broader questions still need to be answered: What is the correct, or at least most reasonable, rate of adherence to a given prescribing guideline? Are face-to-face interventions (either one on one or in small groups) always necessary to address strongly held incorrect beliefs? What should we consider a "clinically important" improvement for a complex practice change strategy? Can reminder systems that are sometimes effective in correcting errors of omission change more resistant errors of commission? Even if single reminders are effective, is there a point of multiple reminder fatigue and diminishing clinical returns? Lastly, are advanced CDSS safe and effective, and, if so, are they worth the time, effort, and opportunity costs necessary to implement and use them? The most recent, rigorous research questions the efficacy, efficiency, and safety of health information technology mandated by the economic stimulus of 2008 [87].

Practice settings may also influence the choice of interventions to be evaluated. For example, organized systems of clinicians (e.g., medical groups, independent practice associations, integrated delivery systems) may be conducive to participatory approaches in which practicing physicians, and possibly patients, work with a facilitator/educator to explore current practices and barriers to change, and then develop or modify practice guidelines, along with methods to measure guideline adherence. These group meetings also serve as vehicles for active learning and begin to converge with the strategy of group detailing described earlier. In addition, we believe more attention needs to be paid to the study of changing the behavior of busy physicians in community practice. Many successful strategies may not be transferable from a university hospital to a busy ambulatory clinic.

Most of the studies we reviewed were designed to assess only whether an intervention changed behavior; few studies have undertaken formal cost–benefit analyses [22,93]. Several formal economic analyses of academic detailing trials demonstrated that the interventions led to a net saving from the societal or organizational perspective [83,93,139]. This is a clear illustration of what Eddy described as "getting more for less," the potential to improve quality and reduce costs simultaneously [140]. There are still relatively few controlled studies that compare the costs and benefits of alternative approaches to improving practice, and little has been published on when (or at what cost) it is reasonable to introduce an intervention to change physicians' practice [93].

Although we know that prescribing problems exist, we still know surprisingly little about their prevalence or determinants. This paucity of

data is all the more remarkable considering that three-quarters of all physician visits end in the prescription of a drug. In a study of more than 30 000 hospital admissions, drug-related complications were common and estimated to account for 19% of all adverse events, and almost one-third of adverse drug events were preventable [141]. Less is known about the ambulatory setting [142–144], with estimates of preventable adverse drug events ranging from 11% to 28%. One retrospective analysis of a New England malpractice insurance carrier observed that 6% of all malpractice claims were related to adverse drug events, and that half of these claims occurred for events in the outpatient setting [144]. These investigators also estimated that three-quarters of these adverse drug events were preventable, and that most medication errors occurred as a result of system deficiencies (e.g., inadequate monitoring) or performance errors (e.g., wrong drug or wrong dose). Like most descriptions of adverse drug events, these studies documented only errors of commission; the extent of omission (e.g., underuse of effective therapies) has been extremely understudied (see Chapter 24).

Finally, studies examining the economic outcomes of interventions as well as studies that include patient-reported outcomes would advance the field. While policy-induced reductions in the use of essential medications have been associated with adverse events [145], few analogous patient outcomes studies exist in the literature on interventions to improve prescribing. Important effects of medications on many health outcomes have been demonstrated in clinical trials; therefore, it is reasonable to hypothesize that more appropriate use of some medications could reduce morbidity and mortality, increase patient functioning, and improve quality of life. Whether improved prescribing is a surrogate measure, or an outcome that directly leads to improved health outcomes, it remains a critically important area for future study.

References

1 Morgan PP. Are physicians learning from what they read in journals? *Can Med Assoc J* 1985; **133**(4): 263.

2 Coleman K, Katz E, Menzel H. *Medical Innovation: A Diffusion Study*. Indianapolis, IN: Bobbs-Merrill, 1966.

3 Rogers E. *Diffusion of Innovations*. New York: Free Press, 1995.

4 Kotler P, Roberto E. *Social Marketing: Strategies for Changing Public Behaviour*. New York: Free Press, 1989.

5 Fox R, Mazmanian P, Putnam RW. *Changing and Learning in the Lives of Physicians*. New York: Praeger, 1989.

6 Greco PJ, Eisenberg JM. Changing physicians' practices. *N Engl J Med* 1993; **329**(17): 1271–3.

7 Davis D, Evans M, Jadad A, *et al*. The case for knowledge translation: shortening the journey from evidence to effect. *BMJ* 2003; **327**(7405): 33–5.

8 Cabana MD, Rand CS, Powe NR, *et al*. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999; **282**(15): 1458–65.

9 Godin G, Belanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. *Implement Sci* 2008; **3**: 36.

10 Eccles M, McColl E, Steen N, *et al*. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ* 2002; **325**(7370): 941.

11 Lu CY, Ross-Degnan D, Soumerai SB, Pearson SA. Interventions designed to improve the quality and efficiency of medication use in

managed care: a critical review of the literature – 2001–2007. *BMC Health Serv Res* 2008; **8**: 75.

12 Soumerai SB, Avorn J. Principles of educational outreach ("academic detailing") to improve clinical decision making. *JAMA* 1990; **263**(4): 549–56.

13 Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. *N Engl J Med* 2007; **357**(6): 608–13.

14 Hennessy S, Soumerai SB, Lipton HL, Strom BL. Drug utilization review. In Strom BL, ed., *Pharmacoepidemiology*, 4th edn. Chichester: John Wiley & Sons, 2005, pp. 439–54.

15 Lipton HL, Byrns PJ, Soumerai SB, Chrischilles EA. Pharmacists as agents of change for rational drug therapy. *Int J Technol Assess Health Care* 1995; **11**(3): 485–508.

16 Beney J, Bero LA, Bond C. Expanding the roles of outpatient pharmacists: effects on health services utilisation, costs, and patient outcomes. *Cochrane Database Syst Rev.* 2000; **3**: CD000336. doi:10.1002/14651858.CD000336

17 Soumerai SB, Lipton HL. Computer-based drug-utilization review – risk, benefit, or boondoggle? *N Engl J Med* 1995; **332**(24): 1641–5.

18 Hennessy S, Bilker WB, Zhou L, *et al*. Retrospective drug utilization review, prescribing errors, and clinical outcomes. *JAMA* 2003; **290**(11): 1494–9.

19 Forsetlund L, Bjorndal A, Rashidian A, *et al*. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2009; **2**: CD003030. doi:10.1002/14651858.CD003030.pub2

20 Grimshaw JM, Shirran L, Thomas R, *et al*. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001; **39**(8 Suppl 2): 2–45.

21 Balas EA, Boren SA, Brown GD, Ewigman BG, Mitchell JA, Perkoff GT. Effect of physician profiling on utilization. Meta-analysis of randomized clinical trials. *J Gen Intern Med* 1996; **11**(10): 584–90.

22 Pearson SD, Schneider EC, Kleinman KP, Coltin KL, Singer JA. The impact of pay-for-performance on health care quality in Massachusetts, 2001–2003. *Health Aff (Millwood)* 2008; **27**(4): 1167–76.

23 Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood)* 2005; **24**(1): 138–50.

24 Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is $1 billion a year a good investment? *Am J Manag Care* 2007; **13**(12): 670–6.

25 Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care* 2006; **15**(6): 433–6.

26 Huskamp HA, Deverka PA, Epstein AM, Epstein RS, McGuigan KA, Frank RG. The effect of incentive-based formularies on prescription-drug utilization and spending. *N Engl J Med* 2003; **349**(23): 2224–32.

27 Bloor K, Freemantle N. Lessons from international experience in controlling pharmaceutical expenditure. II: Influencing doctors. *BMJ* 1996; **312**(7045): 1525–7.

28 Soumerai SB, Ross-Degnan D. Prescribing budgets: economic, clinical, and ethical perspectives. *Aust Prescriber* 1997; **20**: 28–9.

29 Soumerai SB, Ross-Degnan D, Fortess EE, Abelson J. A critical analysis of studies of state drug reimbursement policies: research in need of discipline. *Milbank Q* 1993; **71**(2): 217–52.

30 Soumerai SB, McLaughlin TJ, Ross-Degnan D, Casteris CS, Bollini P. Effects of a limiting Medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia. *N Engl J Med* 1994; **331**(10): 650–5.

31 Smalley WE, Griffin MR, Fought RL, Sullivan L, Ray WA. Effect of a prior-authorization

requirement on the use of nonsteroidal antiinflammatory drugs by Medicaid patients. *N Engl J Med* 1995; **332**(24): 1612–17.

32 Tamblyn R, Laprise R, Hanley JA, *et al*. Adverse events associated with prescription drug cost-sharing among poor and elderly persons. *JAMA* 2001; **285**(4): 421–9.

33 Cochrane Effective Practice and Organization of Care Group. About us. 2017. http://epoc.cochrane.org (accessed October 2017).

34 Spinewine A, Schmader KE, Barber N, *et al*. Appropriate prescribing in elderly people: how well can it be measured and optimised? *Lancet* 2007; **370**(9582): 173–84.

35 Majumdar SR, Soumerai SB. Why most interventions to improve physician prescribing do not seem to work. *CMAJ* 2003; **169**(1): 30–31.

36 McGlynn EA, Asch SM, Adams J, *et al*. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003; **348**(26): 2635–45.

37 Lee TH. A broader concept of medical errors. *N Engl J Med* 2002; **347**(24): 1965–7.

38 Chang AK, Bijur PE, Esses D, Barnaby DP, Baer J. Effect of a single dose of oral opioid and nonopioid analgesics on acute extremity pain in the emergency department: a randomized clinical trial. *JAMA* 2017; **318**(17): 1661–7.

39 Friedman BW, Dym AA, Davitt M, *et al*. Naproxen with cyclobenzaprine, oxycodone/acetaminophen, or placebo for treating acute low back pain: a randomized clinical trial. *JAMA* 2015; **314**(15): 1572–80.

40 Gonzales R, Steiner JF, Lum A, Barrett PH, Jr. Decreasing antibiotic use in ambulatory practice: impact of a multidimensional intervention on the treatment of uncomplicated acute bronchitis in adults. *JAMA* 1999; **281**(16): 1512–19.

41 Gurwitz JH, Field TS, Rochon P, *et al*. Effect of computerized provider order entry with clinical decision support on adverse drug events in the long-term care setting. *J Am Geriatr Soc* 2008; **56**(12): 2225–33.

42 Gurwitz JH, Field TS, Harrold LR, *et al*. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003; **289**(9): 1107–16.

43 Berlowitz DR, Ash AS, Hickey EC, *et al*. Inadequate management of blood pressure in a hypertensive population. *N Engl J Med* 1998; **339**(27): 1957–63.

44 Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**(5): 630–8.

45 Majumdar SR, McAlister FA, Soumerai SB. Synergy between publication and promotion: comparing adoption of new evidence in Canada and the United States. *Am J Med* 2003; **115**(6): 467–72.

46 Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment restrictions for prescription drugs under Medicaid. Effects on therapy, cost, and equity. *N Engl J Med* 1987; **317**(9): 550–6.

47 Soumerai SB, Ross-Degnan D, Avorn J, McLaughlin T, Choodnovskiy I. Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. *N Engl J Med* 1991; **325**(15): 1072–7.

48 McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. *N Engl J Med* 1976; **295**(24): 1351–5.

49 Kravitz RL, Epstein RM, Feldman MD, *et al*. Influence of patients' requests for direct-to-consumer advertised antidepressants: a randomized controlled trial. *JAMA* 2005; **293**(16): 1995–2002.

50 Schwartz RK, Soumerai SB, Avorn J. Physician motivations for nonscientific drug prescribing. *Soc Sci Med* 1989; **28**(6): 577–82.

51 Greer AL. The state of the art versus the state of the science. The diffusion of new medical technologies into practice. *Int J Technol Assess Health Care* 1988; **4**(1): 5–26.

52 Majumdar SR, McAlister FA, Furberg CD. From knowledge to practice in chronic

cardiovascular disease: a long and winding road. *J Am Coll Cardiol* 2004; **43**(10): 1738–42.

53 Gilbert JR, Light JR, Mosteller F. Assessing social innovations: an empirical base for policy. In Bennett CA, Lumsdaine AA, eds., *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*. New York: Academic Press, 1975, pp. 39–193.

54 Cochrane Effective Practice and Organization of Care Group. What study designs should be included in an EPOC review and what should they be called? 2017. http://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/what_study_designs_should_be_included_in_an_epoc_review.pdf (accessed November 2017).

55 Cook TD, Campbell DT. *Quasi-Experimentation: Designs and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin, 1979.

56 Shojania KG, Duncan BW, McDonald KM, Wachter RM. Safe but sound: patient safety meets evidence-based medicine. *JAMA* 2002; **288**(4): 508–13.

57 Chaudhry B, Wang J, Wu S, *et al*. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006; **144**(10): 742–52.

58 Black AD, Car J, Pagliari C, *et al*. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011; **8**(1): e1000387.

59 Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989; **67**(2): 268–317.

60 Soumerai SB, Ross-Degnan D, Gortmaker S, Avorn J. Withdrawing payment for nonscientific drug therapy. Intended and unexpected effects of a large-scale natural experiment. *JAMA* 1990; **263**(6): 831–9.

61 Groves RE. Therapeutic drug-use review for the Florida Medicaid program. *Am J Hosp Pharm* 1985; **42**(2): 316–19.

62 Soumerai SB, Starr D, Majumdar SR. How do you know which health care effectiveness research you can trust? A guide to study design for the perplexed. *Prev Chronic Dis* 2015; **12**: E101.

63 Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1981; **114**(6): 906–14.

64 Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001; **20**(3): 391–9.

65 Freemantle N, Eccles M, Wood J, *et al*. A randomized trial of Evidence-based OutReach (EBOR): rationale and design. *Control Clin Trials* 1999; **20**(5): 479–92.

66 Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. *J Gen Intern Med* 1992; **7**(6): 623–9.

67 Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1996.

68 McCullagh P, Nelder JA. *Generalized Linear Models*, 2nd edn. London: Chapman & Hall, 1989.

69 Soumerai SB, McLaughlin TJ, Gurwitz JH, *et al*. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA* 1998; **279**(17): 1358–63.

70 Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006; **6**: 54.

71 Leapfrog Group. Where you get your health care matters. 2017. http://www.leapfroggroup.org (accessed October 2017).

72 Han YY, Carcillo JA, Venkataraman ST, *et al*. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005; **116**(6): 1506–12.

73 Koppel R, Metlay JP, Cohen A, *et al*. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005; **293**(10): 1197–203.

74 Peterson ED, Roe MT, Mulgund J, *et al*. Association between hospital process performance and outcomes among patients with acute coronary syndromes. *JAMA* 2006; **295**(16): 1912–20.

75 Kucher N, Koo S, Quiroz R, *et al*. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *N Engl J Med* 2005; **352**(10): 969–77.

76 Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. *N Engl J Med* 1996; **335**(13): 966–70.

77 Green LW, Kreuter MW. *Health Promotion Planning: An Educational and Environmental Approach*. Mountain View, CA: HJ Kaiser Foundation, 1991.

78 Farmer AP, Legare F, Turcot L, *et al*. Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2008; **3**: CD004398. doi:10.1002/14651858.CD004398.pub2

79 Zwarenstein M, Grimshaw JM, Presseau J, *et al*. Printed educational messages fail to increase use of thiazides as first-line medication for hypertension in primary care: a cluster randomized controlled trial [ISRCTN72772651]. *Implement Sci* 2016; **11**(1): 124.

80 Grudniewicz A, Kealy R, Rodseth RN, Hamid J, Rudoler D, Straus SE. What is the effectiveness of printed educational materials on primary care physician knowledge, behaviour, and patient outcomes: a systematic review and meta-analyses. *Implement Sci* 2015; **10**: 164.

81 Giguere A, Legare F, Grimshaw J, *et al*. Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012; **10**: CD004398. doi:10.1002/14651858.CD004398.pub3

82 Schaffner W, Ray WA, Federspiel CF, Miller WO. Improving antibiotic prescribing in office practice: a controlled trial of three educational methods. *JAMA* 1983; **250**(13): 1728–32.

83 Avorn J, Soumerai SB. Improving drug-therapy decisions through educational outreach. A randomized controlled trial of academically based "detailing." *N Engl J Med* 1983; **308**(24): 1457–63.

84 Soumerai SB, Avorn J. Economic and policy analysis of university-based drug "detailing." *Med Care* 1986; **24**(4): 313–31.

85 Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J. Opinion leaders vs audit and feedback to implement practice guidelines. Delivery after previous Cesarean section. *JAMA* 1991; **265**(17): 2202–7.

86 Grimshaw JM, Thomas RE, MacLennan G, *et al*. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004; **8**(6): iii–iv, 1–72.

87 Naci H, Soumerai SB. History bias, study design, and the unfulfilled promise of pay-for-performance policies in health care. *Prev Chronic Dis* 2016; **13**: E82.

88 Flodgren G, Parmelli E, Doumit G, *et al*. Local opinion leaders: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2011; **8**: CD000125. doi:10.1002/14651858.CD000125.pub4

89 Flodgren F, Parmelli E, Doumit G, *et al*. Local opinion leaders: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007; **1**: CD000125. doi:10.1002/14651858.CD000125.pub3

90 Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing physician behavior: what works? *Am J Manag Care* 2015; **21**(1): 75–84.

91 Johnson MJ, May CR. Promoting professional behaviour change in healthcare: what interventions work, and why? A theory-led overview of systematic reviews. *BMJ Open* 2015; **5**(9): e008592.

92 O'Brien MA, Rogers S, Jamtvedt G, *et al*. Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007; **4**: CD000409. doi:10.1002/14651858.CD000409.pub2

93 Mason J, Freemantle N, Nazareth I, Eccles M, Haines A, Drummond M. When is it cost-effective to change the behavior of health professionals? *JAMA* 2001; **286**(23): 2988–92.

94 Inui TS, Yourtee EL, Williamson JW. Improved outcomes in hypertension after physician tutorials. A controlled trial. *Ann Intern Med* 1976; **84**(6): 646–51.

95 Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. *N Engl J Med* 1982; **306**(9): 511–15.

96 Fordis M, King JE, Ballantyne CM, *et al*. Comparison of the instructional efficacy of Internet-based CME with live interactive CME workshops: a randomized controlled trial. *JAMA* 2005; **294**(9): 1043–51.

97 Diwan VK, Wahlstrom R, Tomson G, Beermann B, Sterky G, Eriksson B. Effects of "group detailing" on the prescribing of lipid-lowering drugs: a randomized controlled trial in Swedish primary care. *J Clin Epidemiol* 1995; **48**(5): 705–11.

98 Simon SR, Majumdar SR, Prosser LA, *et al*. Group versus individual academic detailing to improve the use of antihypertensive medications in primary care: a cluster-randomized controlled trial. *Am J Med* 2005; **118**(5): 521–8.

99 Stross JK, Bole GG. Evaluation of a continuing education program in rheumatoid arthritis. *Arthritis Rheum* 1980; **23**(7): 846–9.

100 Lomas J. Teaching old (and not so old) docs new tricks: effective ways to implement research findings. In Dunn EV, Norton PG, Stewart M, Bass MJ, Tudives F, eds., *Disseminating New Knowledge and Having an Impact on Practice*. New York: Sage, 1994, pp. 1–18.

101 Majumdar SR, Tsuyuki RT, McAlister FA. Impact of opinion leader-endorsed evidence summaries on the quality of prescribing for patients with cardiovascular disease: a randomized controlled trial. *Am Heart J* 2007; **153**(1): e21–e28.

102 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012; **6**: CD000259. doi:10.1002/14651858.CD000259.pub3

103 O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in Australia. *BMJ* 1999; **318**(7182): 507–11.

104 Glassman PA, Belperio P, Simon B, Lanto A, Lee M. Exposure to automated drug alerts over time: effects on clinicians' knowledge and perceptions. *Med Care* 2006; **44**(3): 250–6.

105 Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA* 2001; **285**(22): 2871–9.

106 Hersey CO, Porter DK, Brelau D, Cohen DI. Influence of simple computerized feedback on prescription changes in an ambulatory clinic. A randomized clinical trial. *Med Care* 1986; **24**: 472–81.

107 Gehlbach SH, Wilkinson WE, Hammond WE, *et al*. Improving drug prescribing in a primary care practice. *Med Care* 1984; **22**(3): 193–201.

108 Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med* 2003; **163**(12): 1409–16.

109 Hershey CO, Goldberg HI, Cohen DI. The effect of computerized feedback coupled with a newsletter upon outpatient prescribing charges. A randomized controlled trial. *Med Care* 1988; **26**(1): 88–94.

110 Koepsell TD, Gurtel AL, Diehr PH, *et al*. The Seattle evaluation of computerized drug profiles: effects on prescribing practices and resource use. *Am J Public Health* 1983; **73**(8): 850–5.

111 Wallerstedt SM, Kindblom JM, Nylen K, Samuelsson O, Strandell A. Medication reviews for nursing home residents to reduce mortality and hospitalization: systematic review and meta-analysis. *Br J Clin Pharmacol* 2014; **78**(3): 488–97.

112 Christensen M, Lundh A. Medication review in hospitalised patients to reduce morbidity and mortality. *Cochrane Database Syst Rev* 2016; **2**: CD008986. doi:10.1002/14651858.CD008986.pub3

113 Shojania KG, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. Effect of point-of-care computer reminders on physician behaviour: a systematic review. *CMAJ* 2010; **182**(5): E216–E225.

114 Arditi C, Rege-Walther M, Wyatt JC, Durieux P, Burnand B. Computer-generated reminders delivered on paper to healthcare professionals: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2012; **12**: CD001175. doi:10.1002/14651858

115 Arditi C, Rege-Walther M, Durieux P, Burnand B. Computer-generated reminders delivered on paper to healthcare professionals: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2017; **7**: CD001175. doi:10.1002/14651858.CD001175.pub4

116 Isaac T, Weissman JS, Davis RB, *et al*. Overrides of medication alerts in ambulatory care. *Arch Intern Med* 2009; **169**(3): 305–11.

117 Nanji KC, Slight SP, Seger DL, *et al*. Overrides of medication-related clinical decision support alerts in outpatients. *J Am Med Inform Assoc* 2014; **21**(3): 487–91.

118 Dexter PR, Perkins S, Overhage JM, Maharry K, Kohler RB, McDonald CJ. A computerized reminder system to increase the use of preventive care for hospitalized patients. *N Engl J Med* 2001; **345**(13): 965–70.

119 Bates DW, Gawande AA. Improving safety with information technology. *N Engl J Med* 2003; **348**(25): 2526–34.

120 Moja L, Kwag KH, Lytras T, *et al*. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. *Am J Public Health* 2014; **104**(12): e12–e22.

121 Bayoumi I, Al Balas M, Handler SM, Dolovich L, Hutchison B, Holbrook A. The effectiveness of computerized drug-lab alerts: a systematic review and meta-analysis. *Int J Med Inform* 2014; **83**(6): 406–15.

122 Gillaizeau F, Chan E, Trinquart L, *et al*. Computerized advice on drug dosage to improve prescribing practice. *Cochrane Database Syst Rev* 2013; **11**: CD002894. doi:10.1002/ 14651858

123 Sahota N, Lloyd R, Ramakrishna A, *et al*. Computerized clinical decision support systems for acute care management: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. *Implement Sci* 2011; **6**: 91.

124 Soumerai SB, Koppel R. Avoiding expensive and consequential health care decisions based on weak research designs. *Health Affairs Blog*. 2015. https://www.healthaffairs.org/do/10.1377/hblog20150831.050258/full (accessed May 2019).

125 Garg AX, Adhikari NK, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005; **293**(10): 1223–38.

126 Soumerai SB, Ross-Degnan D, Kahn JS. Effects of professional and media warnings about the association between aspirin use in children and Reye's syndrome. *Milbank Q* 1992; **70**(1): 155–82.

127 Lu CY, Zhang F, Lakoma MD, *et al*. Changes in antidepressant use by young people and

suicidal behavior after FDA warnings and media coverage: quasi-experimental study. *BMJ* 2014; **348**: g3596.

128 Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med* 2009; **361**(4): 368–78.

129 Serumaga B, Ross-Degnan D, Avery AJ, *et al*. Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. *BMJ* 2011; **342**: d108.

130 Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of premier pay for performance on patient outcomes. *N Engl J Med* 2012; **366**(17): 1606–15.

131 Asch DA, Troxel AB, Stewart WF, *et al*. Effect of financial incentives to physicians, patients, or both on lipid levels: a randomized clinical trial. *JAMA* 2015; **314**(18): 1926–35.

132 Lu CY, Law MR, Soumerai SB, *et al*. Impact of prior authorization on the use and costs of lipid-lowering medications among Michigan and Indiana dual enrollees in Medicaid and Medicare: results of a longitudinal, population-based study. *Clin Ther* 2011; **33**(1): 135–44.

133 Law MR, Lu CY, Soumerai SB, *et al*. Impact of two Medicaid prior-authorization policies on antihypertensive use and costs among Michigan and Indiana residents dually enrolled in Medicaid and Medicare: results of a longitudinal, population-based study. *Clin Ther* 2010; **32**(4): 729–41; discussion 716.

134 Soumerai SB, Zhang F, Ross-Degnan D, *et al*. Use of atypical antipsychotic drugs for schizophrenia in Maine Medicaid following a policy change. *Health Aff (Millwood)* 2008; **27**(3): w185–w195.

135 Zhang Y, Adams AS, Ross-Degnan D, Zhang F, Soumerai SB. Effects of prior authorization on medication discontinuation among Medicaid beneficiaries with bipolar disorder. *Psychiatr Serv* 2009; **60**(4): 520–7.

136 Adams AS, Zhang F, LeCates RF, *et al*. Prior authorization for antidepressants in Medicaid: effects among disabled dual enrollees. *Arch Intern Med* 2009; **169**(8): 750–6.

137 Lu CY, Adams AS, Ross-Degnan D, *et al*. Association between prior authorization for medications and health service use by Medicaid patients with bipolar disorder. *Psychiatr Serv* 2011; **62**(2): 186–93.

138 Lu CY, Soumerai SB, Ross-Degnan D, Zhang F, Adams AS. Unintended impacts of a Medicaid prior authorization policy on access to medications for bipolar illness. *Med Care* 2010; **48**(1): 4–9.

139 Simon SR, Rodriguez HP, Majumdar SR, *et al*. Economic analysis of a randomized trial of academic detailing interventions to improve use of antihypertensive medications. *J Clin Hypertens (Greenwich)* 2007; **9**(1): 15–20.

140 Eddy DM. Clinical decision making: from theory to practice. Rationing resources while improving quality. How to get more for less. *JAMA* 1994; **272**(10): 817–24.

141 Bates DW, Cullen DJ, Laird N, *et al*. Incidence of adverse drug events and potential adverse drug events: implications for prevention. ADE Prevention Study Group. *JAMA* 1995; **274**(1): 29–34.

142 Gandhi TK, Weingart SN, Borus J, *et al*. Adverse drug events in ambulatory care. *N Engl J Med* 2003; **348**(16): 1556–64.

143 Tierney WM. Adverse outpatient drug events – a problem and an opportunity. *N Engl J Med* 2003; **348**(16): 1587–9.

144 Rothschild JM, Federico FA, Gandhi TK, Kaushal R, Williams DH, Bates DW. Analysis of medication-related malpractice claims: causes, preventability, and costs. *Arch Intern Med* 2002; **162**(21): 2414–20.

145 Majumdar SR, Soumerai SB. The unhealthy state of health policy research. *Health Aff (Millwood)* 2009; **28**(5): w900–w908.

20

Pharmacoepidemiologic Studies of Vaccine Safety

Robert T. Chen[1], Jason M. Glanz[2,3], and Tom T. Shimabukuro[4]

[1] *Brighton Collaboration, Task Force for Global Health, Decatur, GA, USA*
[2] *Institute for Health Research, Kaiser Permanente Colorado, Denver, CO, USA*
[3] *Department of Epidemiology, Colorado School of Public Health, Denver, CO, USA*
[4] *Immunization Safety Office, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA*

Vaccines are among the most cost-effective and prevalent primary public health interventions [1,2]. Where immunizations are widely practiced, rates of targeted vaccine-preventable disease (VPD) have declined considerably [3,4]. However, no vaccine is perfectly safe or effective [5]. With high rates of vaccinations and a low incidence of VPD, adverse events following immunizations (AEFI) are understandably of concern, and have received increasing attention [6–10]. Unfortunately, this concern has often negatively affected the stability of immunization programs [11]. For example, questions about the safety of pertussis vaccine in Japan and elsewhere during the 1970s reduced the coverage rate for this vaccine, resulting in the resurgence of pertussis [12]. Similar concerns in the US led to lawsuits, substantial vaccine price increases, and loss of vaccine manufacturers [13], and were a potential deterrent to the development of new vaccines [14]. In the 1990s, concerns about the safety of mercury-based thimerosal preservative used in vaccines [15,16] and the safety of vaccines for anthrax [17] and smallpox [18] affected the stability of US civilian and military immunization programs, respectively. In the UK, a case series report of autism following measles–mumps–rubella (MMR) vaccination in a small number of patients (n = 12; subsequently retracted) precipitated widespread vaccine safety concerns, leading to reduced MMR vaccination rates and subsequent measles outbreaks [19,20]. Similarly, vaccine safety concerns have affected public acceptance of hepatitis B vaccine in France [21], oral polio vaccine (OPV) in Nigeria [22,23], human papilloma virus (HPV) vaccine [24], and 2009 influenza A (H1N1) pandemic (pH1N1) vaccine in several countries [25,26].

More recently, parents have expressed concerns about the safety of the immunization schedule as a whole, worrying that children receive too many vaccines at a very young age and that the recommended immunization schedule overwhelms the immune system [27–29]. These sentiments cause parents to delay or refuse certain or all vaccines for their children, contributing to the adoption of

The authors wish to thank Dr. Komal Narwaney and Dr. David McClure for their contributions to section on "Study Design and Analytic Methods," and Dr. Frank Destefano and Dr. Deb Gust for their review. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

alternative immunization schedules in several high-income countries.

In the early 1990s, the US Institute of Medicine (IOM) – later renamed the National Academy of Medicine (NAM) – reviewed vaccine safety knowledge and research capacity [30,31] and noted it had been limited by (i) inadequate understanding of biologic mechanisms underlying AEFI; (ii) insufficient or inconsistent information from case reports and case series; (iii) inadequate size or length of follow-up of many population-based epidemiologic studies; (iv) limitations of existing surveillance systems in providing persuasive evidence of causation; and (v) few experimental studies published relative to the total number of epidemiologic studies published. IOM/NAM concluded that, "if research capacity and accomplishments [are] not improved, future reviews of vaccine safety [will be] similarly handicapped." Many research and knowledge gaps continue to be identified in each IOM/NAM review of specific immunization safety controversies since 2001, ranging from autism to unexpected infant deaths [17,32–40]. Pharmacoepidemiology has played a vital role since in providing the scientific methods for assessing vaccine safety in the US [41], Europe [42,43], and globally [10,44]. In this chapter, we discuss the major differences in how epidemiology is applied to vaccines versus other pharmaceutical products, giving consideration to both policy and methodology.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Policy Issues

Vaccines share many characteristics with other pharmaceuticals, such as their phased development and licensure, but differ fundamentally in other ways [41]. Understanding these differences is important to appreciate the policy context of vaccine safety and the role of pharmacoepidemiology. Vaccines, for example, are biologic products that are inherently more complex than most small-molecule drugs – both constituent components and the production process [45,46]. Each component of the vaccine formulation – the immunogen, conjugated protein [47], preservative [15], adjuvant [48,49], stabilizer [50,51], diluent [52], and other excipients – has its respective safety considerations (e.g., sourcing, production, quality assurance, safety profile), individually as well as combined [53]. Programmatic errors such as mixing up vaccine vials and unsafe injection practices can also be a concern, especially with poverty [52,54] and in the context of conflicts [55].

A higher standard of safety is also expected of vaccines. In contrast to other pharmaceuticals administered mostly to persons who are ill for curative or therapeutic purposes, vaccines are generally given to healthy people to prevent disease. Tolerance of adverse reactions to products given to healthy people – especially healthy babies – is especially low. This lower risk tolerance for vaccines translates into a need to detect and investigate the possible causes of much rarer events than would be acceptable for other pharmaceuticals. Events that occur at ~$1/10^5$–$1/10^6$ doses administered, such as acute encephalopathy after whole-cell pertussis vaccine [30], Guillain-Barré syndrome (GBS) after swine [56] or 2009 pH1N1 [57] influenza vaccines, and OPV-associated paralytic polio [58], are of concern for vaccines. In contrast, side effects are essentially universal for cancer chemotherapy, and gastrointestinal side effects are very common (10–30%) among people on high-dose aspirin therapy [59].

The cost and the difficulty of studying events increase with their rarity, however (see Chapter 3). Furthermore, conclusions from epidemiologic studies of rare events are less definitive. Attributable risks at about $1/10^5$–$1/10^6$ are considered to be at the margin of resolution for epidemiologic methods [30,60]. The whole-cell pertussis vaccine safety concern in the late 1970s

[12,30] illustrated this challenge. All British children from 2 to 35 months of age hospitalized for several neurologic illnesses over three years (n = 1167) were enrolled in a very large case–control study; however, the finding of a statistically significant association between vaccine and permanent brain damage was based on a small number of exposed cases with wide confidence intervals [61]. Whether or not this study finding was valid, it generated much controversy in and out of the courts [30,62]. Recent advances suggest rare *de novo* genetic mutation and not vaccine as the likely cause [63].

Despite considerably more robust data linking GBS with the 1976–1977 swine influenza vaccine [56], subsequent controversy resulted in a court-ordered independent reexamination of the data and ultimately to a partial repeat of the study, confirming the initial findings [64]. Robust results from two studies on rhesus-rotavirus vaccine and intussusception [65,66] were also challenged [67], until the risk was also shown with the second-generation vaccine, albeit one log rarer [68].

Perhaps not surprisingly, but adding to the confusion, much of the published literature on vaccine safety (and resultant media scares) historically has been in the form of case reports and case series (e.g., a subsequently retracted *Lancet* article alleging links between measles vaccination and autism [19]) rather than controlled studies with adequate power [30,31]. This problem is being ameliorated with the advent of carefully controlled large-linked database studies in several, mostly high-income countries [69,70].

A higher standard of safety is also required for vaccines because of the large number of people who are exposed, some of whom are compelled to be vaccinated by law or regulation for public health reasons [71]. Public health authorities have implemented such requirements because many VPDs (e.g., measles) are highly infectious. When a high proportion of the population is immunized, it creates "herd immunity," so that some of the remaining unimmunized people will still be protected [72]. Without such mandates, a "tragedy of the commons" may occur where high vaccine coverage is reached and the individual benefit/risk ratio diverges from the societal benefit/risk ratio [73,74]. Persons may try to avoid the risks of vaccination while being protected by the herd immunity resulting from others being vaccinated. However, this "commons" provided by herd immunity may disappear if too many people avoid vaccination, with the resulting tragedy that outbreaks return [75], as was experienced in the UK with both pertussis [12] and measles [20], and in the US with measles [76]. A similar policy consideration occurs for some mandatory military vaccinations like those against anthrax [17] and smallpox [18], where a higher vaccine reaction rate may be accepted in exchange for force readiness.

Because of the need for almost universal exposure to many vaccines, the medical maxim "first do no harm" applies even more in public health than in clinical medicine (where decisions usually affect fewer people). Inadequately inactivated polio vaccine (IPV) was administered to about 400 000 people in the "Cutter Incident," resulting in 260 cases of polio [77]. The following incidents and others [78] have fortuitously not resulted in any documented harm to date. Nevertheless, they highlight the importance of ensuring the safety of a relatively universal human-directed "exposure" like immunizations.

- Polio vaccine contaminated by simian virus 40 may have been received by millions of people during the 1950s [35].
- Some vaccines may have contained gelatin stabilizers produced in cattle infected with bovine spongiform encephalopathy [79].
- Some US infants might have been exposed to doses of ethylmercury from thimerosal preservatives in vaccines, exceeding some federal safety guidelines established for ingestion of methylmercury, another form of organic mercury [15].
- Two of the new rotavirus vaccines were contaminated by porcine circovirus [80].

These concerns are the basis for strict regulatory control and other oversight of vaccines by national regulatory authorities such as the US Food and Drug Administration (FDA), European Medicines Agency (EMA), and World Health Organization (WHO) [45,81,82]. Modern technology will continue to improve the ability to detect contaminants in vaccines and influence regulatory decisions during manufacturing [83]; postlicensure monitoring will continue to be important should such findings raise safety concerns.

Very high standards of accuracy and timeliness for results are needed because vaccine safety studies have extremely narrow margins for error. Unlike many classes of drugs for which other effective therapy may be substituted, vaccines generally have few alternative strains or types (OPV and IPV being the best-known exceptions). The decision to withdraw a vaccine [84] or switch between strains may also have wide ramifications [85]. In 1992, the UK authorities stopped procuring mumps vaccines with Urabe strain after studies suggested a high rate of vaccine-associated meningitis [86]. The manufacturers subsequently withdrew this product worldwide, leaving countries without an alternative vaccine if the Urabe strain was their sole licensed mumps vaccine [87]. Safety concerns led to the withdrawal in the early 2000s of what were then the only licensed vaccines against rotavirus [65,66] and Lyme disease [88], rendering these vaccines unavailable anywhere. Establishing associations of AEFI with vaccines and timely measurement of the attributable risk are critical in placing AEFI in the proper benefit/risk perspective. An erroneous association or attributable risk, especially with misinformed media or websites, can undermine confidence in a vaccine and have negative consequences for vaccine acceptance and disease incidence [20]. On the other hand, denials of association despite accumulating evidence can erode public confidence and compromise vaccination programs. For example, public dismay with delayed action on Urabe mumps vaccine-associated aseptic meningitis in Japan forced the Ministry of Health to rescind compulsory school MMR vaccination requirements in 1993 [89].

Because many vaccinations are mandated for public health reasons and no vaccine is perfectly safe, several countries have established compensation programs for people who may have been injured by vaccination. Accurate assessment of whether AEFI can be caused by specific vaccines is essential to a fair and efficient vaccine injury compensation program [90]. In the US, for example, the Vaccine Injury Table (VIT) contains the vaccines, adverse events, onset intervals, and other criteria after which no-fault decisions are made in favor of the claimants. Periodic revisions of the VIT are necessary to reflect the best scientific information on associations between vaccines and adverse events, especially following the introduction of new vaccines [91].

Finally, recommendations for the use of vaccines represent a dynamic balancing of risks and benefits. Vaccine safety monitoring is necessary to weigh this balance accurately. In the face of a meningococcal B epidemic in New Zealand, it was prudent to fast-track the licensure of a new vaccine with limited prelicensure safety data but assurances of good postmarketing surveillance [92]. Even though the second-generation rotavirus vaccine still increases the risk of intussusception, its lower attributable risk (1–6 excess cases/10^5 vaccinees) has been accepted relative to the major reduction in rotavirus disease [68]. When the target diseases are close to eradication, high vaccine complication rates relative to that of the target wild-type disease may lead to discontinuation or decreased use of the vaccine, as was done with smallpox vaccine [93]. Another example was the shift from live OPV to IPV to control OPV-associated paralytic polio and circulation of OPV-derived polio virus [94]. There may be a tradeoff between safety and cost, however. Some countries continue to use Urabe mumps vaccine despite its higher risk for

aseptic meningitis, after the manufacturer lowered the price [95].

With the fears of bioterrorism and importation of previously eliminated VPDs like measles and polio, stopping immunizations and allowing formation of gaps in herd immunity no longer seems advisable [96]. Almost all immunizations will likely be needed indefinitely, with their attendant adverse reactions and potential for loss of public confidence. Because of the success of immunizations in the near elimination of their target diseases, most healthcare providers (let alone parents) have not ever seen a case of some wild-type VPDs. Each future generation must therefore be convinced of the need to be immunized, despite an increasingly remote experience of wild-type disease but contemporary fear of AEFI.

Research on vaccine safety – while applying pharmacoepidemiologic principles – can help to distinguish true vaccine reactions from coincidental events, estimate their attributable risk [56,61,66,97–99], identify risk factors that may constitute valid contraindications [100,101], and, if the pathophysiologic mechanism becomes known, develop safer vaccines [102]. Equally importantly, such research demonstrates a commitment to reducing disease from all causes, vaccine preventable and vaccine induced, and may help to maintain public confidence in immunizations and the credibility of immunization programs.

Clinical Issues

Vaccines, like other pharmaceutical products, undergo extensive safety and efficacy evaluations in the laboratory, in animals, and in phased human clinical trials before licensure (see Chapter 1). Phase I trials usually include a small number of subjects, and can only detect extremely common AEFI. Phase II trials generally enroll hundreds of subjects. When they are carefully coordinated, important conclusions such as the relationship between concentration of antigen, number of vaccine components, formulation technique, effect of successive doses, and profile of common reactions can be drawn from such trials [103]. Such studies can also affect the choice of the candidate vaccine for Phase III [45].

Sample sizes for Phase III vaccine trials generally range between 5000 and 10 000 people, which is larger than most drug trials. *In extremis*, >600 000 schoolchildren were enrolled in the famous Francis field trial of inactivated Salk poliovirus vaccine [104]. To help rule out links with a rarer outcome like intussusceptions (background rate ~5 per 1000 infant years), the second-generation rotavirus vaccine trials enrolled ~70 000 infants [68]. Traditionally, however, sample sizes for Phase III vaccine trials have been based primarily on efficacy considerations; inferences on safety are drawn to the extent possible based on the sample size (~100–100 000) and the duration of observation (often less than 30 days) [45,105]. This usually means that observations of the common local and systemic reactions (e.g., injection site swelling, fever, fussiness) have been possible. Because of the experimental randomized, double-blind, placebo-controlled design of clinical trials, inferences on the causal relationship of an AEFI with the vaccine are relatively straightforward [30,31]. Brazilian investigators also used such a design to compare the risk of aseptic meningitis among three mumps vaccine strains [106]. However, study of rarer outcomes, vaccine exposures (e.g., specific permutation of vaccine antigens administered simultaneously), or subpopulations is usually only practical postlicensure.

Better standardization of safety evaluations in prelicensure clinical trials is needed so that safety data across trials and vaccines can be compared (see also "Classifications and Case Definitions"). In the Phase III trials for infant diphtheria, tetanus, acellular pertussis (DTaP) vaccine, a standard case definition was developed for efficacy, but ironically not for safety – the main reason for the development of DTaP.

For example, definitions of high fever across trials varied by the temperature (39.5 °C versus 40.5 °C), the mode of measurement (oral versus rectal), and time after vaccination measured (48 versus 72 hours) [107]. However, for rarer events, it may be difficult to have standardized assessments across cultures and health systems, as illustrated in the Swedish and Italian trials, in which major differences were detected in rates of hypotonic-hyporesponsive episodes after the same whole-cell pertussis vaccine [108].

The finding of delayed excess mortality in some recipients of high-titer measles vaccine in developing countries [109], now believed by some researchers to be due to a change in vaccine sequence [110] or nonspecific effects of vaccinations [111], led to a call for increasing the current limited duration of follow-up for AEFI in most trials [112,113]. This need was reinforced recently when, six years after the trial, recipients of a new dengue vaccine not previously infected by the virus developed more severe disease with subsequent dengue infection [114].

Many other new vaccines under development (e.g., malaria, tuberculosis) or recently licensed (e.g., rotavirus) are targeted for initial introduction in resource-limited settings. Both pre- and postlicensure safety studies will be needed, for longer follow-up periods in settings where the pharmacovigilance infrastructure is limited or nonexistent [115,116].

Ideally, pharmacogenomics (see Chapter 34) and biobanking can be integrated into prelicensure trials (continuing through to postlicensure) to begin improving our understanding of the biologic/genetic basis for why some persons underrespond and others overrespond to an immunization with respect to immunogenicity and reactogenicity, as we shift from "one size fits all" to more personalized vaccinology [117,118]. Historically, the strategy to deal with vaccine recipients with insufficient immune response was straightforward, consisting of a multidose schedule. Those with overly vigorous reactions on the other hand were more problematic, and potentially at risk of being unfairly labeled "anti-vaccine" if they questioned the safety of receipt of subsequent doses.

Despite over 200 years since Jenner pioneered the smallpox vaccine, the medical science of diagnosing, managing, preventing, or treating rare, serious vaccine reactions remain relatively rudimentary. The reasons are multifold and the challenges are as much logistic as scientific. Modern medicine cannot make progress on rare disorders like leukemia (or rare serious vaccine reactions) by relying on primary care providers alone. Instead, tertiary subspecialties with an adequate referral base and research funds (e.g., hematology/oncology) are needed. With the exception of certain regions in Italy [119], Australia [120], and seven civilian Clinical Immunization Safety Assessment (CISA) project sites [121] and four military Regional Vaccine Safety Hubs [122] in the US, a similar well-organized, well-identified subspecialty infrastructure has been missing for the study of rare vaccine reactions in most countries. Such centers can also potentially play a role for studying newly hypothesized AEFI syndromes [19,123]. The diversity of vaccine exposures (active/passive, live/inactivated, single/combined, etc.), combined with the range of adverse event outcomes (in essence the entire medical textbook, including some not yet defined), means that the new subspecialty will need to play a "case manager" role of drawing upon other subspecialty expertise as needed. However, most importantly, such CISA-type centers could potentially prevent outliers (for example, based on genetic susceptibility) in reactogenicity response from becoming anti-vaccine by recruiting them into mutually beneficial opportunities to improve our scientific understanding and prevent vaccine reactions [124,125] (e.g., a trial of a safer booster dose in children with extensive limb swelling after pertussis vaccination) [227].

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

Signal Detection

Because biologics like vaccines are generally manufactured in living systems rather than via chemical synthesis like drugs, variation in rate of adverse reactions by manufacturer or even lot might be expected. Surveillance systems need to detect such potential anomalies in the expected number and type of AEFI in a timely manner. Some factors make identification of true safety signals difficult. Many vaccines are administered early in life, at a time when the baseline risk is constantly changing and may be affected by other infant events. Furthermore, by definition, if vaccination rates are high, most people with adverse medical events will have had a history of vaccination. Distinguishing causal from coincidental events on a case-by-case basis is rarely possible (see Chapter 33), particularly for events where the pathophysiologic mechanisms are not known, regardless of vaccination. Since many vaccinations are administered to individuals either simultaneously or as a combination vaccine, unless the number of people who also receive that exact permutation of vaccine exposures (including manufacturer and lot number) is known so that AEFI rates can be calculated, it may be difficult or impossible to know if an aberration has occurred. Similarly, when vaccine coverage rates are high and multiple vaccinations are administered concurrently, it can be difficult to disentangle the individual effects of each component, since simultaneous vaccination patterns are likely to be uniform across the population.

Unlike many public health surveillance systems, which focus on either a single exposure (e.g., lead) or a single disease outcome (e.g., measles), vaccine safety surveillance systems need to examine multiple exposures – for instance, different vaccine antigens (frequently administered in combination or simultaneously), manufacturers, and lot numbers – and multiple disease outcomes. Until the recent advent of data-mining methods (see Chapter 46), detection of a vaccine safety signal occurred as much due to a persistent patient [126] as due to data analysis [127]. The tradeoff between sensitivity and specificity depends critically on whether the goal of the surveillance is the detection of a previously unknown illness or syndrome (sensitivity > specificity) or tracking a known disease (specificity > sensitivity). Vaccine safety surveillance systems are asked to monitor *both* previously known and previously unknown AEFI in the same system, however [128]. Nevertheless, the goal of early detection of an aberrant cluster of new AEFI remains identical to other pharmacovigilance and public health surveillance systems.

Standard Definitions and Evaluative Protocols

Case definitions can be used at the time of reporting or at the time of analysis to improve specificity. Applying definitions at the time of reporting may reduce the number of reports processed and lower the operating cost [129]. The sensitivity of surveillance may be lower and the difficulty of assessing misclassification greater, however. Alternatively, if the reporting form is open-ended [130], this may increase the sensitivity of surveillance, but only at the cost of sorting through many nonspecific reports. Definitions can then be applied at the time of analysis. Nevertheless, substantial variation in diagnostic work-up and description of events makes *post hoc* classification difficult without additional follow-up information, which in turn is usually costly.

Historically, it was challenging if not impossible to compare and collate vaccine safety data across clinical trials or surveillance systems in a valid manner because of lack of standard case definitions. We can advance our scientific knowledge of immunization safety by using a

common vocabulary, particularly helpful in the prelicensure setting where maximizing the yield of safety data may help with limited sample sizes. The Brighton Collaboration (see "Solutions: Classifications and Case Definitions") is addressing this gap [107].

Assessment of Causality

Aside from events like local reaction or anaphylaxis, assessing whether any AEFI was actually caused by vaccine is generally not possible unless a vaccine-specific clinical syndrome (e.g., myopericarditis in healthy young adult recipients of smallpox vaccine [18]), or repeat exposures resulting in the same AEFI (e.g., alopecia and hepatitis B vaccination [126]), or a vaccine-specific laboratory finding (e.g., Urabe mumps vaccine virus isolation [131]) can be identified. Whenever the adverse event can also occur in the absence of vaccination (e.g., seizure), a very large clinical trial or more affordable epidemiologic study is necessary to assess whether vaccinated people are at higher risk than unvaccinated people. As noted earlier, when multiple vaccinations are administered simultaneously, determining whether events are attributable to particular components or one of several combinations is frequently difficult or impossible.

Exposure

Misclassification of exposure status may occur if there is poor documentation of vaccinations. Unlike children of school age where vaccination documentation is often required, ascertaining vaccination status in adults may be particularly difficult. In the US, recent and likely future increases in the number of licensed vaccines, the relative lack of combination vaccines, plus, historically, the high mobility among immunization providers (up to 25% annually) because of changes in health insurance plans have led to a potentially confusing maze of vaccination history misclassifications [132].

For example, even though only the acellular pertussis vaccine is available in the US, AEFI reports of the old whole-cell pertussis "DTP" vaccine continue to be received – presumably due to errors in recording by immunization nurses. An infant may have started their immunization series with one provider who uses DTaP combination vaccine from manufacturer A, but switched to another provider to complete the series with DTaP combination vaccine from manufacturer B. Add in the complexity of whether other vaccines like those against polio or hepatitis B are administered simultaneously, at different dose series in the schedule, at different ages, or using different lots of vaccine, and the number of permutations of vaccine exposures that need assessment for potential safety concerns quickly escalates. The availability of complete documentation of vaccine exposure on a large cohort of children in the Vaccine Safety Datalink (VSD) project allowed evaluation of the safety of thimerosal preservatives and the immunization schedule via multiple studies [133].

Outcome

Because of the higher standard required for vaccine safety (as discussed previously), events being assessed are frequently rare (e.g., encephalopathy, GBS) and identifying enough cases for a meaningful interpretation of study findings can be a major challenge. Even when technically feasible, a study may be logistically infeasible or the findings likely to be too inconclusive to justify the resources. This was the conclusion of a 1989 IOM/NAM committee that evaluated whether the UK's National Childhood Encephalopathy Study should be replicated in the US [62].

The difficulty in achieving adequate study power is further compounded in assessing rare events in populations less frequently exposed (e.g., early use soon after introduction on the market, vaccines given to travelers, or subpop-

ulations with special indications). This challenge is well illustrated in studies of the potential association between GBS, which occurs at a background rate of about 1 per 100 000 person-years, and various vaccines. Study of GBS after newly introduced meningococcal conjugate vaccine required assembling data from 9 million adolescents [134]. A retrospective study of GBS after the 1992–1994 influenza vaccinations required assessing hospital records of over 20 million people for 2 years [135]. Active GBS case finding among a population of 45 million may have detected an attributable risk of one additional case of GBS per million 2009 pH1N1 vaccinations [57]. Whenever both the rarity of the adverse outcome and the number of exposures limit the ability to assess a small potential increased risk, identifying risk factors of such rare associations imposes an additional (and possibly prohibitive) level of sample size requirements – unless multinational collaborations are organized.

Many AEFI hypothesized to be caused by vaccines have poorly defined etiologies (e.g., encephalopathy [61], GBS [56], chronic fatigue syndrome [136], narcolepsy [137], sudden unexplained infant deaths [138]). Attributing the outcome to vaccination can only be done after all other potential etiologies have been ruled out, and even then causality cannot be certain. Our scientific understanding of some diseases is frequently limited in the absence of vaccination, let alone with vaccination. This poor understanding severely limits clinical and epidemiologic studies of these illnesses. Furthermore, in highly vaccinated populations, risk-interval analyses (where a specific risk/exposure period is assigned) may be the only epidemiologic study design possible (see "Study Designs, Analyses, Confounding, and Bias"). Predicting the onset of illness following an environmental exposure is critical in calculating the biologically plausible risk interval. For certain hypothesized AEFI, there is no known biologic mechanism to allow prediction of the risk interval. Diseases with insidious or delayed onset like autism [19], inflammatory bowel disease [139], and multiple sclerosis [36] do not permit prediction of the risk interval and are therefore also difficult to study.

Study Designs, Analyses, Confounding, and Bias

Analyzing observational studies of vaccine safety poses several methodologic challenges. Traditional epidemiologic study designs, such as the cohort and case–control designs, are limited because a large percentage of the population tends to be vaccinated. This implies that few unvaccinated individuals are available for analysis, and the unvaccinated tend to differ from the vaccinated by several potential confounding variables, including ethnicity, socioeconomic status, use of the healthcare system, and underlying health disorders [140,141].

Another challenge is that serious AEFI are rare. Cohort studies typically require hundreds of thousands or even millions of study subjects to be able to detect an association between vaccination and the suspected adverse event [56,61,64–66,131]. Such studies can be prohibitively expensive, unless all the requisite information is automated and linkable.

A possible alternative to the cohort design is the case–control study design, in which cases are sampled from the source population and compared to a group of randomly selected event-free controls. This design is well suited for rare events, and has been used for several studies of vaccine safety [65,98,142–144]. It is, however, particularly difficult to choose an appropriate control group without introducing selection bias if the study is not population based. Moreover, because childhood vaccines are generally administered on an age schedule and many childhood illnesses that may be potential AEFI are age dependent, age may confound exposure–outcome relations, for example DTP vaccine and febrile seizures or sudden infant death syndrome (SIDS) [145].

Consequently, such factors must be controlled, generally by matching and subsequent adjustment in the statistical analysis.

To help address these limitations, self-controlled study designs have been developed and implemented [146,147]. These designs involve cohorts of vaccinated individuals (risk interval) or analyses where vaccinated cases are compared to themselves (self-controlled case series). Such designs have been shown to be efficient and valid alternatives to the traditional epidemiologic study designs [148–150]. For details on these methods, see "Methodologic Approaches."

More difficult to control are factors leading to delayed vaccination or nonvaccination [140,141]. Such factors (e.g., low socioeconomic status, preceding illness, parental choice) may confound studies of AEFI and lead to underestimates of the true relative risks. The extent of distortion introduced by confounding can be examined as a function of six variables (Table 20.1). Relatively little is known about the nature, frequency, and implications of these variables, however [140].

Table 20.1 Variables determining the extent of bias attributable to confounding in studies of vaccine adverse events (AE).

Variable	Description
S	Risk of AE in unvaccinated children who lack the contraindication*
R	True relative risk of AE associated with vaccination
D	Relative risk of AE associated with the contraindication
C	Proportion of children with the contraindication
V	Proportion vaccinated among children without the contraindication
P	Proportion vaccinated among children with the contraindication

*"Contraindication" used here to mean any factor associated with avoidance or delay of vaccination.
Source: Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1992; **136**(2): 121–35. Reproduced by permission of Oxford University Press.

Currently Available Solutions

Prelicensure

Standardized toxicity grading scales, initially developed to evaluate products treating cancer and AIDS, have been developed for preventive vaccine trials [151]. Their use allows for meaningful interpretation of prelicensure safety data, including rules for stopping the trial, especially when combined with standardized case definitions for adverse outcomes [107,152].

Whenever potentially important safety signals are detected in prelicensure trials (e.g., intussusceptions after rotavirus vaccine [153]), it is critical that they are pursued postlicensure [154]. Given the need for improved understanding of the safety of vaccines administered universally to healthy babies and the methodologic difficulties of assessing safety postlicensure, some researchers have argued that larger experimental trials may be needed to better assess rare but serious vaccine risks [105,155]. This could be done either with larger prelicensure trials, as has been done with antipyretics in children [155–157] and the post-rhesus rotavirus vaccine trials [68], or in some organized stepwise manner postlicensure (e.g., registry of the first million vaccinations), prior to universal recommendation of the vaccine for entire birth cohorts [156]. Even with these measures, separate large-scale, long-term randomized intervention trials would theoretically be the only way to study unforeseen delayed adverse effects [105,158] or nonspecific effects of immunizations [111]. Such trials would have to avoid withholding efficacious vaccines from people in need. Therefore, maximizing both the pre- and postlicensure assessment processes, as discussed in this chapter, remains optimal.

Data and Safety Monitoring Boards (DSMBs) represent an area of potential improvements in the prelicensure process. Currently, such DSMBs are constituted uniquely for each clinical trial. If instead there is greater overlap across prelicensure trials for the same vaccine, the DSMB may have a better ability to oversee the safety data for the experimental vaccine. The Council of International Medical Organizations (CIOMS) has also proposed an internationally harmonized Development Safety Update Report (DSUR) for summarizing the safety experience for a clinical trial (or entire development program). When aligned with the postapproval Periodic Safety Update Report (PSUR) for marketed products, these could be integrated into a single harmonized safety report that would cover a product throughout its life cycle [159].

Furthermore, despite its name, there is currently no requirement for the DSMB to include someone with drug/vaccine safety experience. For vaccine trials, someone with rare disease (versus infectious disease) epidemiology skills, usually fine-tuned from postlicensure safety monitoring experience, should be considered for the DSMB.

Another area of potential improvement is the method used to determine the likelihood of a causal relationship of an AEFI with the experimental exposure (e.g., a new vaccine; see Chapter 33). Traditionally, the principal investigator of a clinical trial makes an assessment of the causal relationship; this procedure is difficult to standardize and is prone to bias [160]. In an era of increasing automation of medical records and sophistication of methods for detecting nonrandom clusters or elevated rates, similar approaches to assessing prelicensure safety data are required. Finally, there is a need to improve clinical trial infrastructure in resource-limited settings for assessing the safety and efficacy of various preventive and therapeutic products for poverty-related diseases [115,161,162].

With the biotechnology revolution, new candidate viral vector vaccines targeting challenging VPDs like Ebola and HIV are entering human clinical trials. The Brighton Collaboration Viral Vector Vaccines Safety Working Group (V3SWG) was formed to improve the ability to anticipate potential safety issues and meaningfully assess or interpret safety data, thereby facilitating greater public acceptance of a vaccine when licensed [163]. The V3SWG has a standardized template describing the key characteristics of a novel vaccine vector to facilitate the scientific discourse among key stakeholders and increase the transparency and comparability of its risk/benefit information [164]. The V3SWG is also developing standardized guidance on critical issues such as potential risk of recombination between vaccine vector and wild-type virus strains [165].

Postlicensure

Passive Surveillance or Spontaneous Reporting Systems

Informal or formal passive surveillance or spontaneous reporting systems (SRS) have been the cornerstone of most national vaccine safety monitoring systems because of their simplicity and relatively low cost [41,166]. As these terms imply, AEFI reports are generally voluntarily submitted to public health or regulatory agencies, and in some cases to vaccine manufacturers who in turn report, and the agencies passively receive these reports, rather than actively collecting information on AEFI. The national reporting of AEFI can be done through the same reporting channels as those used for other adverse drug reactions [166], as is the practice in many European countries [167] and Japan [168]. Historically, however, few countries have forwarded their AEFI reports to the Uppsala Monitoring Center [169] (see also Chapter 10). An increasing number of countries are collecting safety data specific to vaccinations, either with reporting forms and/or surveillance systems different from the drug safety monitoring systems. These countries include Australia [170],

Brazil [171], Canada [172], China [173], Cuba [174], Denmark [175], India [176], Italy [177], Germany [178], the Netherlands [179], New Zealand [180], Switzerland [181], and the US [182]. Vaccine manufacturers also maintain SRS for their products [183], which are usually forwarded subsequently to appropriate national regulatory authorities [45].

Because of their importance in infectious disease control, a significant proportion of vaccines in many countries is purchased or administered by national public health authorities. For example, in the US the public sector (federal, state, and local governments) purchases over half of the childhood vaccines administered. In many developing countries, the Ministry of Health in conjunction with the WHO's Expanded Program for Immunizations (EPI) administers almost all vaccines. Potential AEFI commonly are first reported by the healthcare providers who administered the vaccine. In many countries, such healthcare providers also participate in public health surveillance for other diseases. Public health authorities (e.g., the US Centers for Disease Control and Prevention, CDC) therefore commonly lead or collaborate with the vaccine licensure and regulatory agency (e.g., the FDA) in developing and administering AEFI reporting systems. A similar model for harmonization and avoiding duplication is followed in Canada and six European countries [167], and is highly recommended for other countries [184].

The US Experience

The US National Childhood Vaccine Injury Act of 1986 mandated for the first time that healthcare providers report certain AEFI [71,185]. The Vaccine Adverse Event Reporting System (VAERS) was implemented jointly by the CDC and FDA in 1990 to provide a unified national system for collection of reports of AEFI, including but not limited to those mandated for reporting [130,182]. To increase sensitivity, the VAERS form is designed to permit narrative descriptions of AEFI. All people, including patients or their parents and not just healthcare professionals, are permitted to report to VAERS, especially on medically important events. Vaccine manufacturers are required to report any AEFI that comes to their attention.

In 2016, 22% of US VAERS reports were submitted by patients, parents, or other unidentified sources, 27% by healthcare providers, and 51% by vaccine manufacturers; in recent years leading up to and including 2016, VAERS received around 40 000 US reports annually (CDC, unpublished data). There are no restrictions set on interval between vaccination and onset of illness or requirements that a patient must have received medical care in order for the event to be reported. Reports are accepted without judgment on whether the vaccine caused the AEFI. Although reporting is encouraged as soon as possible after the AEFI, there is no time limit for reporting.

Web-based reporting became available in 2002; experience to date shows it to be more complete and timely [186] and it was therefore heavily used during the 2003 US smallpox [187] and 2009 pH1N1 vaccination campaigns [188]. In 2017, the CDC and FDA implemented a revised VAERS reporting form and reporting process, called VAERS 2.0, which, along with the FDA eVAERS initiative using the FDA Electronic Submissions Gateway, allows for fully electronic reporting for the public and healthcare providers, and vaccine manufacturers [189]. Future potential developments include (i) enriched passive surveillance in VAERS using clinical decision support systems in electronic health record (EHR) systems to facilitate physicians' identification of possible AEFI, automatically populate data elements of an AEFI report, and enable direct electronic reporting, resulting in more accurate, complete, efficient, and timely transmission of VAERS reports [190]; and (ii) incorporating immunization information system (IIS) data to estimate denominators for AEFI reporting rates in VAERS [191]. The latter is

especially important to overcome the problem of interpreting VAERS data in the face of the increasing heterogeneity of vaccine exposures in the US.

Enhancements to VAERS passive surveillance since its inception have included capability for near real-time report review by CDC and FDA physicians and scientists, collaborations with professional medical associations, development of a user-friendly public use data query tool, multiple options for web-based reporting for consumers and healthcare providers, and direct electronic reporting for vaccine manufacturers. There is also more frequent review and dissemination of safety data via publications and reports to advisory committees. "Enhanced passive" surveillance via VAERS has been successfully used to date in safety surveillance for rotavirus [192], yellow fever [193], smallpox vaccine [187], 2009 pH1N1 vaccine [188], tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis (Tdap) vaccine and inactivated influenza vaccine in pregnant women [194,195], and would likely be implemented in any future new vaccine or counter-bioterrorism-related wide-scale vaccination program [196].

Approximately 7% of US VAERS reports are classified as serious (documenting death, life-threatening illness, permanent disability, hospitalization, or prolongation of hospitalization), although the percentages of serious/nonserious reports vary by age group and type of vaccine. A contractor, under CDC and FDA supervision, collects, codes (using the Medical Dictionary for Regulatory Activities or MedDRA; https://www.meddra.org), and enters VAERS data and reports into a database. Trained nurses follow up on AEFI classified as serious to obtain additional medical information and recovery status. The CDC and FDA have access to the VAERS database, the original VAERS reports, and medical records obtained on follow-up, and focus their efforts on analytic tasks of interest to the respective agencies. VAERS data with personal identifiers removed are available to the public as downloadable datasets and through a user-friendly data query tool at https://vaers.hhs.gov/data.html.

Other High-Income Nations' Experiences

Several other countries also have substantial experience with passive surveillance for vaccine safety, including Canada [197] and the Netherlands [198]. The UK and many members of the Commonwealth (e.g., Australia [170]) use the "yellow card" system, where a reporting form is attached to officially issued prescription pads [199]. Korea employs a multipronged approach that includes spontaneous surveillance, rapid response and AEFI investigation, and compensation when appropriate [200]. Data on adverse drug (including vaccine) events from many nations are compiled by the WHO Collaborating Center for International Drug Monitoring in Uppsala (www.who-umc.org), which has also begun a vaccine focus [201].

Various approaches (short of database linkage) have been taken to supplement passive systems to help overcome their methodologic weaknesses. Canada (the Immunization Monitoring Program – Active or IMPACT, operational since 1990 [197], with a new variant for influenza [202]), Australia (the Paediatric Active Enhanced Disease Surveillance or PAEDS system [203], and Singapore [204] have active, pediatric hospital-based surveillance systems that searches all admissions for possible relationships to immunizations. In Canada, an Advisory Committee on Causality Assessment, consisting of a panel of experts, previously reviewed the serious passive reports [205].

Low- and Middle-Income Countries' (LMIC) Experience

As recently as 2012, 65% of WHO member states, including the majority of LMIC, did not have a functional postmarketing monitoring system [206,207]. In response, a comprehensive global manual on surveillance of AEFI was developed by WHO [208] as part of the Global

Vaccine Safety Initiative [209]. The manual provides guidance on (i) the objectives of vaccine and immunization safety surveillance; (ii) the AEFI surveillance system: reporting, investigation, causality assessment and classification of cause-specific AEFI; (iii) understanding vaccine reactions for better decision-making; (iv) the best use of surveillance data; and (v) response processes, including a communication strategy on immunization safety for the public and the media. The introduction of single-use autodisable syringes, "bundled" with vaccine procurement by donors, has significantly reduced the number of correctable programmatic errors like injection site abscesses (due to inadequate sterilization) [210].

As more new vaccines are (or are planned to be) first introduced in LMICs, there is increasing awareness of the need to improve currently inadequate pharmacovigilance systems in these countries [115,211]. The decades-long delay in discovering serious AEFI after yellow fever vaccination [212] and BCG vaccination in human immunodeficiency virus-infected infants [213] further highlights this urgent need. The WHO launched the Global Vaccine Safety Initiative in 2012, addressing many of the needs identified in the Global Vaccine Safety Blueprint [209]. While some progress has been attained despite limited funding, much capacity-building remains [207,211]. For example, some of the challenges with causality assessment of AEFI identified by countries in South East Asia included poor quality of data (e.g., lack of autopsies for deaths), high staff turnover rates, and deficits in AEFI investigation training and timely investigation (due to inadequate financial support for AEFI committees and investigation teams [214].

Classifications and Case Definitions

AEFI can be classified by frequency (common, rare), extent (local, systemic), severity (hospitalization, disability, death), causality (probable, possible, unlikely, etc.), and preventability (intrinsic to vaccine, faulty production, faulty administration). Wilson developed the first classification system with a focus on errors of production (e.g., bacterial, viral, toxin contamination) and administration (e.g., nonsterile apparatus) [5]. In 2012, CIOMS/WHO updated its classification of AEFI (Table 20.2) [215].

The distinction between *vaccine induced* and *vaccine precipitation*, as first clarified for DTP and DT vaccine and infantile spasm [216], has

Table 20.2 Types of adverse event following immunization (AEFI) by cause.

Cause	Definition
Vaccine product-related reaction	An AEFI that is caused or precipitated by a vaccine due to one or more of the inherent properties of the vaccine product
Vaccine quality defect-related reaction	An AEFI that is caused or precipitated by a vaccine that is due to one or more quality defects[1] of the vaccine product including its administration device as provided by the manufacturer
Immunization error-related reaction	An AEFI that is caused by inappropriate[2] vaccine handling, prescribing or administration and thus by its nature is preventable
Immunization anxiety-related reaction	An AEFI arising from anxiety about the immunization
Coincidental event	An AEFI that is caused by something other than the vaccine product, immunization error or immunization anxiety

Notes: [1] Any deviation of the vaccine product as manufactured from its set quality specifications; [2] Usage (handling, prescribing, and administration) other than what is licensed and recommended in a given jurisdiction based on scientific evidence or expert recommendations.
Source: Council for International Organizations of Medical Sciences. Definition and application of terms for vaccine pharmacovigilance: report of CIOMS/WHO Working Group on Vaccine Pharmacovigilance. Geneva: CIOMS, 2012. Reproduced by permission of CIOMS.

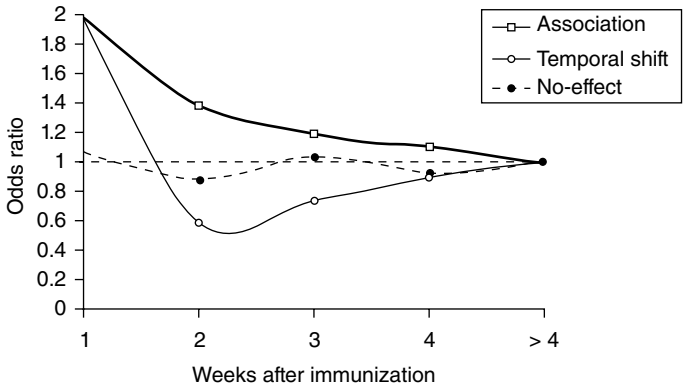


Figure 20.1 Three theoretical models of the temporal relationship between immunization and an adverse effect: (i) Association: the risk exceeds 1 at all time windows postimmunization; (ii) temporal shift: the risk exceeds 1 initially but then falls below 1, but coming back to 1 eventually, such that the area under the curve above and below 1 is similar; and (iii) no effect: the risk stays around 1. *Source:* Goodman M, Lamm SH, Bellman MH. Temporal relationship modeling: DTP or DT immunizations and infantile spasms. Vaccine 1998; **16**: 225–31. Reproduced with permission of Elsevier.

been useful because vaccine precipitation does not result in excess vaccine-attributable risk over time, whereas vaccine induced does (Figure 20.1).

The Brighton Collaboration was established in 2000 as an international voluntary effort to enhance vaccine safety science; it focused initially on the development, evaluation, and dissemination of standardized case definitions of AEFI [107]. Global workgroups of experts are convened to develop case definitions that are then peer reviewed. The Brighton case definitions for each AEFI are arrayed by the level of evidence presented (insufficient, low, intermediate, and highest); therefore, they can also be used in settings with a range of resources (e.g., from prelicensure trials to postlicensure surveillance, or from LMIC to high-income country settings). Over 50 Brighton case definitions are now available for use at www.brightoncollaboration.org, including many related to safety assessment of immunization in pregnancy [217].

Adding the Brighton case definition into the Canadian national reporting form in 2009 increased the proportion of seizure cases meeting the case definition three- to sixfold [172]. Alternatively, in a more open reporting system like VAERS, these definitions can be applied to reports to develop a case series for further investigation [218,219]. Real progress in implementation of similar standards across national boundaries is being realized with the advent of the International Conference on Harmonization (ICH) [220] and the Brighton Collaboration [107]. For example, the wide use of Brighton case definition for studies of GBS after the 2009 pH1N1 pandemic influenza vaccine facilitated meta-analyses and cross comparisons [57].

Standardized Clinical Assessment Protocols and Centers

There has been an increasing awareness that the utility of SRS and vaccine injury programs [221] as potential registries of rare AEFI and the immunization safety infrastructure can be usefully augmented by tertiary clinical centers with expertise in vaccinology and vaccine safety. The US initiated its Department of Defense Vaccine Healthcare Centers Network [124] and the CDC's civilian Clinical Immunization Safety Assessment (CISA) Project [121] to take advantage of this opportunity. These programs bring

together infectious disease epidemiologists, immunologists, neurologists, dermatologists, and other subspecialists from multiple participating sites as needed for various tasks [32]. Among these tasks is the standardized assessment of complex clinical cases of AEFI in individual patients to improve our scientific understanding of the pathophysiology and risk factors of the reaction, and to provide guidance on future vaccination for patients who have experienced AEFI [121,205,222–225]. New understanding of the human genome, pharmacogenomics, and immunology may now make it possible for us to better understand rare or newly recognized reactions for several vaccines (see also Chapter 34) [63,117].

Through these centers, standardized assessment protocols can be developed to examine patients with similar AEFI to see if they constitute a rare or a previously unrecognized clinical syndrome. If so, a case definition can be developed that permits identification of cases for follow-up validation studies examining the potential role of vaccination in causing this syndrome [226].

For patients who have had AEFI that generate concern but do not contraindicate completion of a vaccine series or further vaccination, such as hypotonic-hyporesponsive episodes [222], extensive limb swelling after acellular pertussis vaccination [227], and sterile abscesses following aluminum adjuvant-containing vaccines [228], the standardized clinical assessment centers, such as CISA, can provide assessment and guidance for the management of subsequent vaccinations.

Finally, standardized clinical assessment centers can provide regional referral and advice services – with the opportunity to follow up and document compliance with the advice provided and the outcomes, so that the rare experience can be added to our scientific knowledge. Ultimately, many AEFI diagnostic or management protocols can be made available on the internet for other clinicians to use (and to provide a mechanism for them to contribute their experience) [229]. Both the development and application of standardized case definitions and standardized evaluation of clinical syndromes play a "hypothesis-strengthening" role, intermediate between hypothesis generation and hypothesis testing.

Assessment of Causality

The formal process of assessing causality in the association of an adverse event and an exposure (e.g., vaccine) is complex and can be considered in terms of the answers to three questions: Can It? Did It? and Will It? [230]. The answer to *Can It?* (i.e., the potential for a causal link) was the focus of the IOM/NAM reviews [30,39,231]. It is usually based on population-level inferences drawn from epidemiologic studies and the following considerations: (i) strength of association, (ii) analytic bias, (iii) biologic gradient/dose–response, (iv) statistical significance, (v) consistency, and (vi) biologic plausibility/coherence [232].

For individual case reports, the *Did It?* question is more relevant. If the answer is yes, then *Can It?* is also answered in the affirmative. It is natural to suspect a vaccine to be the cause when an AEFI occurs in temporal association following vaccination. To base causal inference purely on temporal association, however, is to fall for the logical fallacy of *post hoc ergo propter hoc* ("after this, therefore because of this") [31]. Information useful for assessing causality in individual case reports includes (i) previous general experience with the vaccine (e.g., duration of licensure, number of vaccinees, whether similar events have been observed among other vaccinees or nonvaccinees, existence of animal models to test the vaccine as a cause); (ii) alternative etiologies; (iii) biologic plausibility; (iv) individual characteristics of the vaccinee that may increase the risk of the AEFI; (v) timing of events; (vi) characteristics of the event (e.g., laboratory findings); and (vii) rechallenge [233,234] (see also Chapter 33).

When a vaccine *can* cause an AEFI, *Will It?* refers to the probability that an individual will experience the event, or, for populations, the proportion that will experience the event as a result of vaccination (i.e., the attributable risk fraction). These data are critical for developing valid precautions and contraindications for the individuals and benefit/risk policy decisions for the population. *Will It?* is usually very difficult to answer, however, as it can only be addressed based on epidemiologic studies [31]. Furthermore, the sample sizes of such studies may be large enough to establish whether the vaccine can cause a given event, but yet inadequate to stratify by subgroups to examine risk factors that can help delineate potential contraindications [138].

Specific AEFI may be considered to be caused by a specific vaccine if the event is associated with a unique laboratory finding, and/or a very specific clinical outcome. For example, Urabe mumps vaccine virus was implicated as a cause of aseptic meningitis because mumps virus was isolated from the cerebrospinal fluid (a normally sterile body site) and was shown to be a vaccine and not wild-type strain by genetic sequencing [131]; or vice versa for wild-type varicella-induced paralysis, which was initially blamed on the vaccine [123]. The detection of IgG antibodies to the stabilizers in vaccine in children with hypersensitivity reactions confirms the etiology [51]. Demonstrations that severe local swelling following tetanus toxoid tended to occur in people with extremely high levels of circulating antitoxin (due to excessive tetanus boosters) support the proposed mechanism of an Arthus reaction [235]. Acute flaccid paralysis, especially shortly after receipt (or contact with a recipient) of OPV, is almost pathognomonic of OPV-associated paralytic polio in countries where wild-type poliovirus is unlikely to be circulating [58]. Similarly, acute myopericarditis in otherwise healthy recent smallpox vaccinees also supports a causal relationship [18]. Causality can sometimes be inferred if a specific and uncommon clinical finding occurs after each vaccination (i.e., challenge–rechallenge), as in cases of alopecia after hepatitis B vaccination [126]. However, unlike in drug safety, dechallenge (disappearance of the adverse event by stopping the medication) is usually not feasible with immunizations.

If the adverse event is known to be associated with the wild-type VPD (e.g., acute arthritis and idiopathic thrombocytopenic purpura [ITP] after rubella), its association with the live, attenuated vaccine at a lesser frequency is not surprising [236]. This relationship is not universal, however, as pregnant women who receive live attenuated rubella vaccine, unlike those exposed to wild-type rubella, have not been shown to have illness compatible with congenital rubella syndrome [237]. Clustering of events in time after vaccination can also suggest causation if "reporting bias" can be ruled out. Such bias may occur since parents and doctors are most likely to link AEFI with vaccinations the shorter the time interval between the two and the more serious the event. Febrile seizures associated with killed bacterial vaccines tend to occur within a day of vaccination, while those due to live viral vaccines are delayed by about a week due to viral replication [97,238]. Onset of GBS after the swine influenza vaccination was delayed by up to six weeks, but clustered at two to three weeks following vaccination, as autoimmune demyelination is a slower process [56]. The pattern of the risk by time since vaccination may suggest that the relationship to vaccination is more one of temporal shift or triggering of an underlying susceptibility (Figure 20.1) [138,216].

Unfortunately, most serious reported AEFI lack these unique features that permit informed inferences on causality. Autism, chronic fatigue syndrome, SIDS, and GBS either have multiple or as yet unknown etiologies. In a highly vaccinated population, it is not surprising that most cases of any adverse event have a history of prior vaccinations. Epidemiologic studies have to be relied upon to ascertain likelihood of association and, if related, the attributable fraction.

Because of these challenges, some vaccine injury compensation programs may simplify their administrative proceedings by making a blanket assumption that all AEFI occurring within particular periods after vaccination are "caused" by the vaccine, irrespective of whether they were truly causal or just coincidental. This, unfortunately, may lead some individuals to imply inaccurately that all such compensated cases are *caused* by vaccinations. Despite these caveats, the timing of the onset interval after vaccination plays a major role in most causality assessment algorithms, as AEFI after live viral vaccines usually occur later than those of killed vaccines [97,238].

In some countries, expert committees of specialists in relevant disciplines (e.g., pediatrics, infectious disease, neurology) review reports. This "global introspection" approach [230] has been used by some countries [200,239] to classify reports of AEFI in gradations of probable association to vaccination (see also Chapter 33).

The CISA Project developed a standardized algorithm to assist in collecting and interpreting data, and to help assess causality after individual AEFI, building on Canadian lessons [239] and review of 2009 pH1N1 AEFI [240]. The classifications were based on the reported symptoms, the interval between vaccination and onset of symptoms, and a set of case definitions. Final classification generated by the process includes four categories in which the event is either (i) consistent; (ii) inconsistent; or (iii) indeterminate with respect to causal association. This algorithm (with a fourth unclassifiable category) has since been adapted by the WHO [241], with refinements proposed [242].

Because the opinions of experts play such a major role in this form of causality assessment, the results are less satisfying than results obtained from rigorously conducted scientific studies. After a review of available approaches, the European Vaccine Adverse Event, Surveillance & Communication (VAESCO) project concluded: "the usefulness of individual causality assessment of AEFI remains to be demonstrated. Well documented cases and proper case definitions may be more important than causality assessment especially for signal detection and evaluation" [233].

Signal Detection

Identifying a potential new vaccine safety problem ("signal") requires a mix of clinical intuition, epidemiologic expertise, the application of statistical data-mining tools, and, frequently, a large increase in vaccine exposure. Unusual clinical features and/or AEFI clustering in time or space may suggest that something may be awry. Traditionally, a signal occurs when an observed number of events exceeds the number of events expected by chance alone for the specific data source (i.e., the background rate). For example, no illness other than GBS was reported more commonly in the second and third weeks than in the first week after swine influenza vaccination, leading to further validation studies [56]. In general, an acceptable type I error rate is set at 5%, with 80% statistical power to detect a signal. Once a signal has been detected, additional methods such as a temporal scan statistic can be used to detect nonrandom clustering of onset intervals

Several recent examples in the US and elsewhere highlight the importance of rapidly identifying and responding to serious AEFI identified following new vaccines or newly reintroduced vaccines. After a prelicensure signal [243], passive reports to VAERS of intussusception among children vaccinated with rhesus rotavirus vaccine were the first postlicensure signal of a problem [192], leading to several studies to verify these findings [65,66]. Similarly, initial reports to VAERS of a previously unrecognized serious yellow fever vaccine–associated viscerotropic disease [101] and neurotropic disease [193] have since been confirmed elsewhere [244] and as early as 1973, in retrospect [245]. Acute myopericarditis has been a relatively unexpected finding among people vaccinated against smallpox in the US for bioterrorism preparedness [18].

Less clinically serious, but important vaccine safety signals nonetheless, have also been detected. Oculorespiratory syndrome was found in association with influenza vaccines from one Canadian manufacturer in one season [246]. Bell's palsy was detected in recipients of a new Swiss intranasal influenza vaccine [98]). While several GBS cases were reported to VAERS after the introduction of tetravalent meningococcal conjugate vaccine in adolescents in the US, subsequent large controlled studies found no association [247]. Febrile seizures in young children were observed more than expected in passive reporting in Australia and the US following administration of two formulations of trivalent inactivated influenza vaccine (TIV) [248]; after evaluation in the VSD project, this signal was verified, with simultaneous administration of another vaccine contributing to the increased risk [249].

Historically, automated screening for signals using SRS reports had been challenging [250], largely because of the inherent methodologic problems of spontaneous reports (see earlier discussion and Chapter 10). For example, automated signal generation will not flag events that are not uniquely coded (e.g., the coding system may lack a specific term for Sjögren's disease or other rare conditions). However, new tools developed for pattern recognition in extremely large databases have increasingly been applied. VAERS is one of the largest databases for rare AEFI in the world, with hundreds of thousands of reports. Because of its continuously increasing size and the need to monitor a large number of vaccine–symptom combinations, there has been a substantial effort to apply various computer-assisted techniques for automated detection of unusual trends and patterns.

Several different data-mining methods (see Chapter 27) that have been evaluated in VAERS to date include empirical Bayesian [251–253], association rule discovery [254], multi-item gamma Poisson shrinkage [255], proportional morbidity distribution [256], and proportional reporting rate ratio [257,258]. No single method appears to be superior [259]. Rational approaches to prioritizing the large numbers of potential signals generated using automated algorithms on large passive AEFI report databases may involve the utilization of complementary approaches, such as data visualization and an array of different data-mining methods (each with pros and cons), where a cumulative higher score might signal cause for greater concern. Ultimately, these methods represent a useful adjunct to, but not a substitute for, traditional methods of scrutinizing spontaneous reports in increasingly complex databases such as VAERS [257].

A near real-time sequential analytic approach, called rapid cycle analysis (RCA), has been developed by the CDC VSD project (see "Automated Large-Linked Databases") to conduct active surveillance of newly licensed and recommended vaccines, existing vaccines with new recommendations or indications, and seasonal influenza vaccine annually [260,261]. RCA takes advantage of the strengths of the VSD, with its ability to gather automated vaccination exposure, outcomes, and medical care utilization data from enrolled members in several integrated healthcare organizations. The process analyzes data weekly, or when a predetermined number of doses of a vaccine have been administered, and uses statistical techniques to account for multiple comparisons and data lags. RCA in the VSD has not only successfully simulated but also detected an observed increase in febrile seizures after the combination measles–mumps–rubella–varicella (MMRV) vaccine [262] and an increase in febrile seizures in young children following inactivated influenza vaccine (later determined to be associated with concomitant 13-valent pneumococcal conjugate vaccination) [249].

The VSD has a covered population of about 12 million persons per year, which, while large, may still encounter difficulties detecting associations between rare exposures and very rare outcomes, or more common exposures or outcomes among a specific subpopulation, such as

pregnant women. In RCA, outcomes under surveillance are prespecified based on prior knowledge of the vaccine safety profile of the product from clinical trials or vaccine safety concerns in general, and tend to be limited in number compared to data mining, where outcomes are not identified *a priori* and are essentially unlimited. Therefore, RCA is more consistent with hypothesis testing rather than hypothesis generation, and is not a true source of *de novo* signals. New information theory approaches may provide a way of detecting previously unexpected associations after vaccination, including data mining in large-linked databases [263]. Until these new procedures are validated, a large national passive surveillance system such as VAERS is still necessary as an early harbinger of potential vaccine safety signals for very rare or unusual events.

Large Immunization Campaigns

Whenever very large numbers of vaccine doses are administered over a short time interval, this can either result in more prominent clusters of AEFI or, by their absence, can demonstrate their safety. Note that this occurs irrespective of whether the vaccine exposure is part of a planned mass immunization campaign or not. For example, the link drawn between hepatitis B vaccine and demyelinating disease in France was due in part to increased vaccinations beyond the intended adolescent age group [264]. Surveillance of AEFI around the time of mass immunization campaigns have been extremely useful in generating signals, either positive (e.g., allergic reaction after dextran-stabilized measles vaccine [51], viscerotropic disease following yellow fever vaccine [212], aseptic meningitis after mumps vaccine [265], GBS with swine influenza vaccine [56], GBS after OPV [266], allergic reactions after Japanese encephalitis vaccine [267], neuropathy after rubella vaccine [268] or absent (e.g., events after meningococcal vaccine [269], GBS after measles [270]. Such signals still require validation, however, since some, after more careful scientific studies, are not confirmed to represent a true association [271,272]. Mass psychogenic illness can plague mass vaccination campaigns, especially among adolescents in school settings [273].

Preparation in advance of mass vaccination campaigns is critical. During mass campaigns with new group A meningococcal conjugate vaccine [269] and during the large vaccination effort for 2009 pH1N1 influenza vaccine in the US and elsewhere [191,274], several systems were put in place to identify signals early. For the latter, in the US, VAERS offered the earliest available data to determine if there was a safety concern. An active GBS case finding project among a population of 45 million was also able to determine rapidly if there was an increased risk of GBS following 2009 pH1N1 vaccination. Both systems had strength in the population size and the rapid review of reports [275,276]. Additionally, near real-time sequential monitoring for multiple prespecified outcomes was performed in the VSD during the 2009 pH1N1 influenza vaccination campaign in the US [277]. Assessing and having background rates for likely AEFI during mass campaigns is also very helpful [56,274,278]. Special registries or studies are needed, however, to monitor the outcome for subpopulations like pregnant women [279], who may need to be vaccinated with limited safety data during such campaigns. Even with planning, unexpected AEFI (e.g., narcolepsy) combined with media attention may create biases that are difficult to sort out via epidemiologic studies [98,280].

Lessons Learned to Date

Lessons have emerged from SRS like VAERS [182,281–284]. Such systems worldwide have successfully detected previously unrecognized reactions and helped to obtain data to evaluate whether AEFI are causally linked to vaccines [18,98,126,192,193,246,248]. VAERS has also successfully served as a source of cases for further investigations of known associations such

as idiopathic thrombocytopenic purpura after MMR [285], anaphylaxis after MMR [286], and syncope after immunization [287]. VAERS has been of great value for answering routine public queries such as "Has adverse event X ever been reported after vaccine Y?" and describing the postlicensure safety profile of new vaccines [288–292]. Additionally, VAERS has been useful for rapidly assessing the safety of vaccines recommended in special populations when limited safety data existed at the time of the recommendation, such as the case of Tdap being recommended during each pregnancy regardless of prior Tdap vaccination history [293]. More recently, VAERS has demonstrated its usefulness in monitoring preventable vaccine administration program errors [294,295].

When denominator data on doses are available from other sources (e.g., net doses distributed, vaccine coverage surveys, immunization registries), VAERS can be used to evaluate changes in reporting rates over time or when new vaccines replace old vaccines. However, reporting rates may be susceptible to biases from media attention, systems enhancement efforts, or other environmental changes that can increase reporting, making comparison over time difficult. In addition, doses distributed in the marketplace do not necessarily equate with doses administered (i.e., influenza vaccine wastage at the end of each influenza season), and it is not possible to estimate reporting rates in special populations (i.e., those with specific medical conditions) based only on information on gross doses distributed. Comparing the proportion of reports for specific events may be helpful to minimize this type of bias. For example, analysis of VAERS data showed that after millions of doses of DTaP had been distributed, the reported rate for serious events like hospitalization and seizures after DTaP in toddlers was one-third that after DTP [256]. Reports to VAERS of OPV-associated paralytic polio disappeared after the shift from OPV to IPV in the US [283]. The proportion of GBS reports following inactivated influenza vaccines over several seasons did not vary, including following 2009 pH1N1 vaccines, even though the reporting rates for GBS were higher following 2009 pH1N1, which was likely due to stimulated reporting from heightened media attention [296]. VAERS is also currently the only surveillance system that covers the entire US population, and the data are available on a relatively timely basis. It is, therefore, the major means available currently to detect possible new, unusual, or extremely rare AEFI, including whether certain lots of vaccines are associated with unusually high rates of AEFI [283], especially when combined with estimates of lot use denominator obtained from statistical models [297].

Data from SRS such as VAERS have helped to inform the potential clinical management [298] of AEFI and to identify potential risk factors for such events, such as advanced age [101] and thymic dysfunction [299] associated with yellow fever vaccine complications, concurrent zoster infection in varicella vaccinees resulting in meningitis [300], personal and family history of convulsions in pertussis vaccinees [100], and factors associated with postvaccinal syncope-related injuries [301]. Conversely, a review of VAERS reports of febrile seizures in young children following inactivated influenza vaccine confirmed that they were clinically similar to typical uncomplicated febrile seizures [248], providing reassurance to clinicians and the public.

The reporting efficiency or sensitivity (i.e., the proportion of total AEFI reported) of an SRS can be estimated if the expected rates of AEFI generated from carefully executed studies are available. A study using this method showed that a higher proportion of serious events like seizures that follow vaccinations are likely to be reported to VAERS (or its predecessor, the Monitoring System for Adverse Events Following Immunizations or MSAEFI) than milder events like rash, or delayed events requiring laboratory assessment, such as thrombocytopenic purpura after MMR vaccination

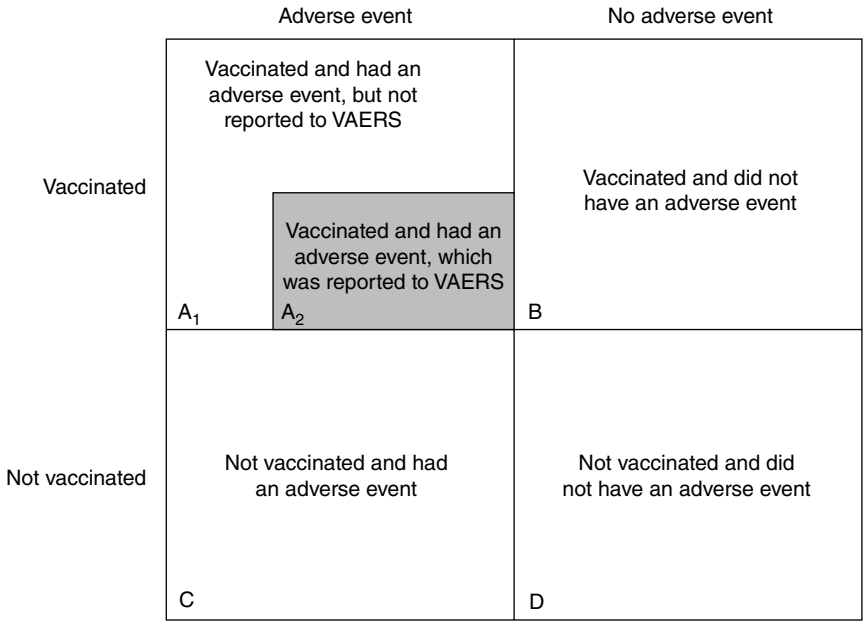
Table 20.3 Reporting efficiencies* for selected outcomes, two passive surveillance systems for adverse event following immunization, US.

		Reporting efficiency(%)		
Adverse event	**Vaccine**	**MSAEFI**	***VAERS* (overall)**	**VAERS (public sector)**
Vaccine-associated polio	Oral polio vaccine (OPV)	72	68	**
Seizures	Diphtheria–tetanus–pertussis (DTP)	42	24	36
Seizures	Measles–mumps–rubella (MMR)	23	37	49
Hypotonic–hyporesponsive episodes	DTP	4	3	4
Rash	MMR	<1	<1	5
Thrombocytopenia	MMR	<1	4	<1

Notes: *Calculated as the ratio of the rates at which adverse events were reported to each passive surveillance system divided by their rates. **Public- and private-sector information is missing on these cases.
MSAEFI, Monitoring System for Adverse Events Following Immunizations; VAERS, Vaccine Adverse Event Reporting System.
Source: Rosenthal S, Chen R. The reporting sensitivities of two passive surveillance systems for vaccine adverse events. *Am J Public Health* 1995; **85**(12): 1706–9. Reproduced by permission of Sheridan Content Solutions (on behalf of The American Public Health Association).

(Table 20.3) [302]. The real incidence of smallpox vaccine myopericarditis in healthy military recruits (presumably averse to admitting chest pain) was ~200-fold higher than passive surveillance [124]. "Capture–recapture" methods, when at least two independent sources are available to ascertain incident AEFI cases in the same population and enough identifying data on the cases are also available to identify individuals ascertained in both dataset sources, can help assess the sensitivity of the reporting systems. Using this method, only an estimated 47% of rhesus rotavirus vaccine–attributable cases of intussusception were reported to VAERS, despite the substantial associated media publicity [303]. Although formal evaluation has been limited, the probability that a serious event reported to VAERS by a healthcare provider has been accurately diagnosed (i.e., predictive value positive) is likely to be high. Of 26 patients reported to VAERS who developed GBS after influenza vaccination during the 1990–1991 season and whose hospital charts were reviewed by an independent panel of neurologists blinded to immunization status, the diagnosis of GBS was confirmed in 22 (85%) [304]. In general, the validity of diagnoses reported to VAERS is highly variable depending on condition.

Despite these uses, SRS for drug and vaccine safety have a number of major methodologic weaknesses (see also Chapter 10) and pitfalls for the unwary in the use of public use datasets [182,284]. Biased and incomplete reporting is inherent to all such SRS and potential safety concerns may be missed [284,302]. Aseptic meningitis associated with the Urabe mumps vaccine strain, for example, was not detected by SRS in most countries about a decade after licensure [87,97,131]. Most importantly, however, the information content of such spontaneous reports represents just cell "a" of a two-by-two table of vaccination versus adverse



A_2 = VAERS database

$$\text{Incidence of AE in vaccinated individuals} = \frac{A_1 + A_2}{(A_1 + A_2)+B}$$

$$\text{Reporting efficiency to VAERS} = \frac{A_2}{A_1 + A_2}$$

$$\text{Incidence of AE in unvaccinated individuals} = \frac{C}{C+D}$$

Figure 20.2 "2 × 2" table necessary for epidemiologic analysis of causality between vaccine and an adverse event following immunization. AE, adverse event; VAERS, Vaccine Adverse Event Reporting System. *Source:* Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS). *Vaccine* 2015; **33**(36): 4398–405. Reproduced with permission of Elsevier.

event (Figure 20.2), and incomplete and biased content at that [305].

Use of data from SRS is further complicated by heterogeneity in reported clinical syndromes, absence of laboratory confirmation of many of the events, and simultaneous vaccinations that make proper attribution of the causal vaccine difficult. MedDRA and WHO Adverse Reaction Terminology (WHO-ART) codes used to classify signs and symptoms documented in AEFI reports do not necessarily represent medically confirmed diagnoses. Since much of "signal detection" relies on specific diagnoses and their coding into databases, new AEFI clinical syndromes may not be "recognized" and analyzed as such until hypothesis-strengthening procedures such as the development of standardized case definitions and/or clinical/laboratory evaluation are undertaken. Researchers in Canada did a series of such studies to characterize then "new" oculorespiratory syndrome after the 2000–2001 influenza vaccination [246], which, in retrospect, probably also occurred in other influenza seasons [306] and other countries with other influenza vaccine manufacturers [307].

Current SRS are also prone to detecting increases in AEFI reporting that are not true increases. Instead, they may be due to an increase

in (i) reporting efficiency; (ii) vaccination coverage; (iii) public awareness about vaccine safety from media attention (i.e., stimulated reporting or awareness bias); or (iv) increases in the incidence of known or unknown etiologies for a particular AEFI. SRS are usually unable to sort out causally related from coincidentally related AEFI because of inherent methodologic weaknesses. For example, an increase in GBS reports to VAERS in 1993–1994 influenza vaccinees compared to 1992–1993 influenza vaccinees was found to be due to improvements in vaccine coverage and increases in GBS background incidence, while the vaccination-associated risk remained unchanged [135]. An increased reporting rate of an AEFI following one hepatitis B vaccine brand compared to another was likely due to differential distribution of brands in the public versus private sectors, which have differential VAERS reporting rates (higher in the public sector) [308]. A signal of venous thromboembolic events in HPV vaccinees in VAERS was probably due to confounding from concurrent use of oral contraceptives [290]. Finally, an approximately two- to threefold increase in 2009 pH1N1 reports to VAERS as compared to the 2009–2010 seasonal influenza vaccine occurred, most likely due to heightened public awareness and enhancements made to VAERS for safety monitoring efforts of the 2009 pH1N1 vaccine [296].

These observations highlight the crude nature of the "signal" generated by VAERS, and the difficulty in ascertaining which vaccine safety concerns warrant further investigation. Not only are there problems with reporting efficiency and potentially biased reporting, but precise denominators for calculating true rates are usually not available. Instead, crude measures such as *doses distributed* must often be used as surrogates for *doses administered*. Because of these difficulties, the requirement for manufacturers to notify FDA whenever they receive an increased number of reports has been dropped [309].

Historically, most (especially resource-limited) countries have relied on SRS alone for postlicensure vaccine safety monitoring. The inadequacy of scientific information on vaccine safety found by the IOM/NAM is related to the methodologic weaknesses inherent to SRS. The establishment of new population-based immunization information systems in which all vaccines administered are entered may provide more timely submission of spontaneous reports, as well as more accurate and specific denominators for doses administered, providing information necessary to calculate more accurate AEFI rates [310].

Clinical Trials

Prelicensure Clinical Trials

To demonstrate that a new vaccine candidate is safer than a previous vaccine, the two products can be compared head to head in a randomized trial, as was done for acellular and whole-cell pertussis vaccine [103]. Alternatively, active surveillance in a large trial can be done to show that the attributable risk for a specific AEFI (e.g., intussusception) was lower for a new rotavirus vaccine, compared to the old one [68]. When another vaccine (vs. placebo) is used for comparison, however, there may be challenges in interpreting the safety signals detected [311]. Separately, when a new AEFI like myopericarditis was recognized after smallpox vaccination, trials of new vaccine candidates using a similar viral vector may require more safety assessment (e.g., electrocardiogram) [312]. The challenges and lessons learned from safety monitoring of recent trials of a new group A meningococcal conjugate vaccine in Africa have been published [313].

Postlicensure Clinical Trials

To optimize vaccine use, clinical trials may be conducted after vaccine licensure to assess the effects of changes in vaccine formulation [314], vaccine strain [106,315], age at vaccination [316], the number and timing of vaccine doses [317], simultaneous administration [318], and

interchangeability of vaccines from different manufacturers [319] on vaccine safety and immunogenicity. The importance of such trials was demonstrated when studies showed an unanticipated differential mortality among recipients of high- and regular-titer measles vaccine in resource-limited countries [109], albeit lower than among unvaccinated children [320]. This finding resulted in a change in recommendations by WHO for the use of such vaccines [321]. The development of automated large-linked databases (LLDB; see later discussion) may permit an improved ability to monitor the safety of such postlicensure changes in vaccine use without necessarily conducting such clinical trials.

Postapproval Surveillance Studies

To improve the ability to detect AEFI that are not detected during prelicensure trials, most recently licensed vaccines in developed countries have undergone formal postapproval surveillance studies on populations with sample sizes of 100 000. These studies have usually used computerized data from cohorts in health maintenance organizations supplemented by diary or telephone interview. These methods were first extensively used after the licensure of polysaccharide and conjugated Hib [322,323], DTaP [324], and varicella vaccines (including multiyear evaluation for disease incidence, herpes zoster, and a pregnancy registry) [325,326]. Postapproval studies are now routine for newly licensed vaccines like MMRV vaccine [99], HPV vaccine [24], and second-generation rotavirus vaccines [68]. Postapproval studies in Mexico and Brazil have found an increased risk of intussusception in the newer rotavirus vaccines, albeit one-tenth that of the first-generation vaccine [68]. Postapproval evaluation has even been extended to less frequently used vaccines, like Japanese encephalitis vaccine [327]. A large postlicensure randomized trial for this vaccine was also completed in China to improve the available data on its short-term safety [328].

Ad Hoc *Epidemiologic Studies*

Historically, *ad hoc* epidemiologic studies have been conducted to assess signals of potential AEFI generated by SRS, the medical literature, or other mechanisms. Traditional analyses of secular trends (ecologic studies), cohort studies, and case–control studies have been used to gather information necessary to measure or compare risks of an AEFI following vaccination with risk in the absence of vaccination. Occasionally, data collected for other study outcomes may be reanalyzed to see if the vaccine was causally related or not. Examples of *ad hoc* follow-up studies to signals of vaccine safety issues are the investigations of poliomyelitis after IPV [77] and OPV [329]; SIDS after DTP vaccination [37,138]; encephalopathy after DTP vaccination [61,62]; meningoencephalitis after mumps vaccination [131]; injection site abscesses postvaccination [330]; intussusception after Rotashield vaccine [65,66,68]; vaccinations and autism [39,70]; GBS after influenza vaccine [56,135,275]; and GBS after meningococcal conjugate vaccine [247]. Many such studies have been compiled and reviewed by the IOM/NAM. While automated LLDB (see next section and Chapters 11–14) provide a more cost-effective and flexible framework for hypothesis testing, *ad hoc* epidemiologic studies may still be needed in settings without automated LLDB [98,246], or where the statistical power of the automated LLDB may be inadequate to answer a question in a timely manner [134,135,304].

Automated Large, Linked Databases

Ad hoc epidemiologic studies of vaccine safety, while potentially informative about vaccine causality, are costly, time consuming, and usually limited to assessment of a single or small number of outcomes. As with drug safety research (see Chapters 11–14), efforts have increasingly turned to record linkage between automated exposure (immunization records in lieu of pharmacy) files and outcome medical files. This is nicely illustrated by comparing the methods of

the first rotavirus vaccine intussusception studies, *ad hoc* [65] versus LLDB [66], the latter being more timely and efficient.

The US Experience

Vaccine Safety Datalink Project

The CDC participated during the late 1980s in two pilot vaccine safety studies using automated LLDBs in Medicaid and Managed Care Organizations (MCO) populations, respectively [331–333]. While validating this approach for vaccine safety studies and providing scientifically rigorous results, these studies were limited by their relatively small sample sizes, inability to prospectively study new hypotheses, and focus on the most severe reactions [236]. These limitations, the constraints of VAERS, and the recognition of the need for improved monitoring of vaccine safety prompted the CDC to initiate the VSD project in 1990 [334]. To help overcome the previously identified shortcomings, the VSD prospectively collects vaccination, medical outcome (e.g., hospital discharge, outpatient visits, emergency room visits, and deaths), and covariate data (e.g., ethnicity and socioeconomic data on birth certificates, census) under joint protocol at multiple MCOs. Selection of staff model prepaid health plans also minimized potential biases for more severe outcomes resulting from data generated from fee-for-service claims, a problem prior to the implementation of diagnosis-related group (DRG) billing [335]. To increase patient confidentiality, the VSD shifted from annual data file submissions from the MCOs for data pooling and analysis at CDC to a distributed network data management model [336]; in parallel, the VSD is also increasing transparency via public access data sharing and external input [261].

Originally, the VSD conducted active surveillance on approximately 500 000 children from birth through 6 years of age (75 000 birth cohort, approximately 2% of the US population in these age groups) [334]. Expansion to eight MCOs (including data on all age groups at three MCOs) was accomplished in 2000 [337]. The VSD focused its initial efforts on examining potential associations between immunizations and 34 serious neurologic, allergic, hematologic, infectious, inflammatory, and metabolic conditions. The VSD is also being used to test new *ad hoc* vaccine safety hypotheses that arise from the medical literature [16,338–340], from VAERS [66,308], from changes in immunization schedules [341,342], or from the introduction of new vaccines [262,343]. In addition, the VSD has conducted influenza vaccine safety studies in which large cohorts of children are screened for evidence of increased medically attended events following vaccination [344]. The size of the VSD population also permits separation of the risks associated with individual vaccines from those associated with vaccine combinations, whether given in the same syringe or simultaneously at different body sites [262,345]. Near real-time surveillance was conducted on several combination vaccines [346,347]. Surveillance on the safety of the influenza vaccine, including subgroups such as pregnant women [342], is ongoing.

When the VSD identifies an AEFI as being associated with a vaccine, data on the incidence rate attributable to the vaccine are available [66,262,340], permitting accurate benefit/risk assessment by both the public and policymakers [348]. Subgroup analyses may permit identification of risk factors for AEFI (or vaccine failures), which may be useful in identifying contraindications to vaccinations [349]. Data from VSD have been useful in calculating background rates of illnesses in the absence of vaccination that can serve as expected rates when comparing rates of vaccine-associated events in an SRS [278]. Also, incidence rates of vaccine-associated AEFI derived from VSD can be used to evaluate the sensitivity of passive reporting systems [303]. The VSD data also aid the Vaccine Injury Compensation Program in determinations of what events should be compensated as vaccine "injuries" [71].

In addition to *ad hoc* epidemiologic studies, an RCA team was formed within the VSD to

conduct near real-time active surveillance on newly licensed vaccines [260,261]. The RCA relies on analytic datasets that are created weekly from the automated MCO data. The weekly analytic datasets are used to investigate potential associations between vaccines and AEFI that are defined *a priori*. Statistical analysis for signal detection is conducted with methods that account for the multiple testing of accumulating data. The RCA team developed a statistical method known as the maximized sequential ratio probability test (MaxSPRT) to detect safety signals in near real time, while accurately accounting for repeated testing of the data [350]. The first application of MaxSPRT was in safety assessment of the newly licensed meningococcal conjugate vaccine in 2005 [351]. RCA methods also were used to detect a twofold increased risk for febrile seizures following MMRV vaccination compared to MMR and varicella (MMR + V) administered separately [262]. This finding precipitated changes in US immunization policy for MMRV and MMR + V in children.

The VSD has some limitations, however. The IOM/NAM, in its 2013 report, recommended that the VSD expand collaborations to include more diversity in its study population. In response, the VSD is collaborating with Denver Health, an integrated safety net healthcare system serving a large proportion of Denver's indigent and minority populations, to integrate its data into the VSD [352]. More importantly, because of the high coverage attained in the MCOs for most vaccines, few nonvaccinated controls are available. Therefore, the VSD often relies on some type of "risk-interval" analysis [331–334] (Box 20.1). The capability of this approach to assess associations between vaccination and AEFI with delayed or insidious onset (e.g., neurodevelopmental or behavioral outcomes) is limited [16]. The VSD also cannot easily assess AEFI that do not result in a healthcare visit, and therefore are not currently captured in existing MCO databases, because they do not result in a healthcare consultation (e.g., fever) [334]. The current VSD is also not large enough to examine modest increased risks of extremely rare events such as GBS after each season's influenza vaccine. Finally, because the VSD relies on observational studies, it may not successfully control for confounding and bias in each analysis [140], and inferences on causality may be limited [353].

Despite these potential shortcomings, the VSD provides an essential, powerful, and relatively cost-effective complement to ongoing evaluations of vaccine safety in the US [69].

Postlicensure Rapid Immunization Safety Monitoring

In 2009, the Department of Health and Human Services initiated the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) program in response to the need to supplement the preexisting active surveillance for the safety of the 2009 pH1N1 influenza vaccine [191]. To ensure its sustainability, PRISM became the immunization safety monitoring component of the FDA's Sentinel project (see Chapter 25), a postmarketing safety surveillance network to actively monitor the safety of medical products [354]. PRISM is a distributed data network that utilizes claims data from four national health insurers and vaccine data from eight immunization registries, and has the largest cohort size in the US for active vaccine safety surveillance, making it possible to study rare AEFI [355]. However, its limitations include updating of the data quarterly only and limited accessibility to medical records. PRISM has conducted several vaccine safety evaluations, including on the safety of monovalent pandemic 2009 pH1N1 influenza vaccine during 2009–2010 [356], febrile seizures after 2010–2011 TIV [357], risk of venous thromboembolism after quadrivalent HPV vaccination [358], and intussusception risk after rotavirus vaccination in US infants [359]. Since the VSD and PRISM have different data sources and populations, they complement each other and serve as a check on the validity of results.

Box 20.1 Example of method for risk-interval analysis of association between a universally recommended three-dose vaccine (with a few unvaccinated people for comparison) and adverse event following immunization.

1) Define "risk interval" for adverse event after vaccination (e.g., 30 days after each dose).
2) Partition observation time for each child in the study into periods within and outside of risk intervals, and sum respectively (e.g., for a child observed for 365 days during which three doses of vaccine were received; total risk interval time = 3×30 person-days = 90 person-days; total nonrisk interval time = 365 – 90 = 275 person-days).

o————————x====————————x====————————x====————————//————————>|

Birth Dose 1 Dose 2 Dose 3 365 days

Add up (i) total risk interval and nonrisk interval observation times for each child in the study (=person-time observed; for mathematical convenience, the example below uses 100 and 1000 person-months of observation), and (ii) adverse events occurring in each time period to complete 2×2 table (for illustration, the example uses 3 and 10 cases):

Adverse event yes		*Person-time observed (months)*	*Incidence rate*
Vaccinated in risk interval yes	3	100	0.03
Vaccinated in risk interval no	10	1000	0.01
Total	13	1100	

Incidence rate adverse event$_{vaccinated}$ = 3/100 = 0.03

Incidence rate adverse event$_{unvaccinated}$ = 10/1000 = 0.01

Relative risk vaccinated: unvaccinated = 0.03/0.01 = 3.0

Probability finding due to chance: <5/100

Conclusion: There is a threefold increase in risk of developing the adverse event within the interval following vaccination.

PRISM is currently evaluating new data-mining techniques such as TreeScan, which can be used to simultaneously evaluate multiple unsuspected but potential AEFI while adjusting for multiple comparisons [360]. If it works, it could be an invaluable tool for signal detection. However, since this would not be able to adjust for all possible confounders, the statistical signals of public health concerns would still need evaluation using carefully designed epidemiologic studies [263].

Other High-Income Nations' Experience

In view of the methodologic and logistical advantages offered by automated LLDB, Denmark [70], the UK [97,131], France [361], Germany [362], and Canada [363] have also developed large automated databases linking immunization registries with medical files. Building on their experience with 2009 pH1N1 influenza vaccine safety surveillance, Europe [42] and Taiwan [274] have converted their systems into large, linked databases for routine vaccinations [138]. The European Accelerated Development of VAccine benefit–risk Collaboration in Europe (ADVANCE) consortium (http://www.advance-vaccines.eu) is a public–private partnership under the Innovative Medicines Initiative [44,364]. The Taiwan system links a National Immunization Information System with the outcomes from a successfully implemented National Health Insurance scheme [138]. Most high-income countries should have similar capabilities if they prioritize vaccine safety.

Low- and Middle-Income Nations' Experience

The first pilot LLDB in an LMIC was established in one province in Vietnam in 2002 [365]. Given that many vaccines against many poverty-related diseases like rotavirus, malaria, and tuberculosis will be introduced first in such countries, there is a need to develop VSD-like infrastructures there, too [44,115,211,366]. Affordable options in the near term include adapting hospital-based surveillance for serious AEFI [162], building on existing demographic surveillance systems like the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) [367], pilot implementation projects associated with the introduction of new vaccines [368], and others [369,370].

Injury Compensation

Many high-income countries in Europe, North America, and Asia have programs to compensate individuals who may have been injured by vaccines [71,371]. A global no-fault program has been proposed to facilitate the development of new vaccines for public health emergencies [90]. Although differences exist across countries, some common principles and procedures guide the administration of compensation programs. These are mostly no-fault programs designed to be fair and as efficient and generous as possible. Vaccinations are in some situations mandatory and often recommended by public health authorities, and most AEFI and injuries are not due to negligence or malice (i.e., they are not foreseeable or preventable at the individual level). Therefore, a formal government-managed program outside of the traditional tort system is advisable to protect individuals, vaccination programs, and vaccine manufacturers. Depending on the type of program, claims of vaccine injury may be adjudicated through legal or administrative processes. Funding sources to compensate individuals include general revenues, manufacturer contributions, or specific taxes on vaccine doses sold. Standard of proof generally follows a civil or modified civil standard, rather than a scientific standard for causality [71,371].

In the US, the National Vaccine Injury Compensation Program (VICP) was authorized by the US National Childhood Vaccine Act of 1986 [71,372]. This no-fault program is administered by the Health Resources and Services Administration of the US Department of Health and Human Services. The US Department of Justice represents the government and cases are heard before the US Court of Federal Claims. Covered vaccines include those routinely recommended by the CDC for use in children; however, there is no age limit and anyone who receives a covered vaccine is covered by the program (e.g., an adult receiving an IPV prior to travel would be covered, since IPV is routinely recommended by the CDC for children). Claims can be conceded by the government if they meet the requirements for a VIT injury; the VIT is designed to streamline adjudication of claims by allowing the government to concede causality in certain circumstances [91]. Claims can also be settled (both injuries listed on the VIT and "off-table" injuries), or cases decided by special masters of the US Court of Federal Claims. Funding for the program is through an excise tax on each dose of covered vaccines sold in the US [71,372].

Injury compensation programs have helped vaccine-injured individuals and their families deal with injury-related financial burdens, improved confidence in vaccinations and immunization programs, and helped stabilize vaccine supply and prices [71]. An aspirational goal should be to extend injury compensation programs to all countries that manage national immunization programs [90,371].

Methodologic Approaches for Observational Epidemiology Studies

Exposures

In countries where vaccinations are required for entry into daycare, kindergarten, schools, and/or colleges, documentation via vaccination

cards or medical records is usually available and of good quality for most infants and children. In the US, documentation of the vaccine type, date of vaccination, manufacturer, lot number, and vaccine provider in a permanent medical record has been required since 1988 for certain routine childhood vaccinations [372]. This requirement, along with improvements in technology, has prompted state and local authorities, in collaboration with healthcare providers and health systems, to capture patient vaccination status, particularly for children, in electronic immunization information systems (formerly known as immunization registries) that often communicate with EHRs, which are increasingly common in higher-income countries [310].

Although vaccination records can be manually retrieved and reviewed for any study design, automated vaccination records greatly ease the logistics of organizing such studies. Whenever sampling is necessary in the design, automated records also ease the selection of samples that are representative. Assessing the accuracy of such automated data is important in any study [373,374]. When people receive their vaccinations from several providers (not uncommon in the US), their exposure status may be misclassified [375]. This error could be minimized if a centralized Immunization Information System (IIS) was implemented to track all vaccinations from birth. Such IISs have been implemented in Australia [376], Canada [377], China [378], Denmark [70], Latin America [379], the UK [380], and the US [310]. Mobile phones and other sources of digital identification may aid adherence to vaccinations and their capture as data [381], including in LMICs [382].

The availability and quality of vaccination records generally decrease as people age, especially beyond school age, where documentation is often required for school attendance. Some vaccines for older people may be administered in settings other than primary healthcare (e.g., tetanus–diphtheria boosters in emergency rooms, hepatitis B vaccinations for healthcare personnel). In addition to review of primary medical records, interviews or a review of data from secondary vaccination sites may therefore be necessary to accurately ascertain exposure status in AEFI studies of these vaccines in older populations. To increase the accuracy of exposure data in a study of adverse reactions to plasma-derived hepatitis B vaccine among Alaskan natives, medical records from the village, the hospital, and the regional public health nurse, in addition to the automated vaccination record, were reviewed [383]. Studies of GBS and 2009 pH1N1 influenza vaccine relied on patient/family interview, hospital medical record, and/or validation with primary care providers for exposure ascertainment [275]. Interestingly, reliance on provider verification may lead to underascertainment of vaccination status, either because of poor record keeping [375] or concerns about liability in vaccine safety studies [135].

Standards are needed to improve the accuracy and efficiency of transfer of vaccine identification information from the vaccine vial to automated or paper immunization records. They include (i) abbreviations for new vaccine antigens and vaccine manufacturers; (ii) peel-off labels; (iii) bar codes; (iv) lot numbers [384]; and (v) presentation of key identifier information on vaccine packaging (as on the nutrition label). The WHO has identified as a priority the development of a vaccine dictionary that will allow differentiation of vaccine formulations from various manufacturers [385].

Outcomes

To ensure both high sensitivity and specificity for an AEFI, a multistep approach is usually required for case ascertainment [338,340,344]. In step one, the automated databases are screened to identify International Classification of Diseases, 10th edition (ICD-10) diagnostic codes for the condition of interest. The ICD-10 codes typically represent medical encounters in the inpatient, outpatient, and emergency department settings. Additional data sources,

such as laboratory and pharmacy files, can also be used to identify potential cases. This initial screening definition tends to be highly sensitive but less specific. After the electronic cases have been identified, a trained abstractor, blinded to vaccination status, often reviews the medical records of the patients. On a standardized data collection form, the abstractor records detailed clinical information on presenting symptoms, sequelae, medications, underlying health conditions, diagnostic test results, and potential confounding variables. For outcomes with insidious onset like multiple sclerosis, multiple dates (e.g., first symptom, first medical visit, first diagnosis) and sources of information (patient recall, medical chart) may also need to be collected [386]. In the last step of the case ascertainment process, clinical experts review the abstracted medical information to determine if patients meet the final study case definition. For difficult diagnoses like GBS, a panel of specialists may also be asked to review the medical records after exposure status has been masked [64,135].

This process minimizes the likelihood of a false negative conclusion (due to bias toward the null) by ensuring that only cases meeting the most specific case definition are included in the analysis. It is also possible, however, that using such a narrowly focused outcome definition may miss broader syndromes or groups of symptoms related to the outcome. Follow-up analyses of rhesus rotavirus vaccine reports to VAERS suggest that intussusception [192] may have been just the "tip of the iceberg" of a broader syndrome that also included bloody stool, vomiting, diarrhea, and abdominal pain [288]. Adverse neurologic outcomes other than GBS were reported among the 1976–1977 and 2009 pH1N1 influenza vaccinees [240,387]. Unfortunately, whether these associations are causal remains unknown and controversial, as formal studies have not been done.

Should the concern be a new previously undescribed syndrome, analyses of existing databases may be inadequate. A study of Gulf War Syndrome and vaccinations relied on a thorough interview of patients meeting a *de novo* complex case definition before linkage with vaccination history [388].

In the context of real-time surveillance, influenza vaccine safety monitoring is hindered by the rate at which LLDBs capture medical encounter data. In the VSD, for example, some of the MCO sites contract with independent hospitals to provide inpatient care. Therefore, there is often a considerable lag between the inpatient encounter and the date at which the encounter (outcome) is captured in the databases. At some sites, the average lag can be as long as four months [389]. For influenza vaccine safety monitoring, the influenza season may be over by the time the outcome data is fully captured, thereby rendering the real-time analysis moot.

Study Design and Analytic Methods

Different analytic strategies are needed depending on how a vaccine is used in the population. For vaccines used infrequently and typically in vaccinees who are generally no different than nonvaccinees (e.g., travel vaccines), comparison between two groups with adequate matching or adjustment is relatively straightforward. For example, in a cohort study, groups of vaccinated and unvaccinated individuals may be matched on several factors such as sex, study site, age, high-risk condition, and calendar time. The cohorts of vaccinated and unvaccinated individuals are then followed forward in time, and the incidence of events in the two groups is compared within predefined exposure windows following vaccination. These exposure windows are defined *a priori* based on the current understanding of the most plausible biologic mechanism, should such an association actually exist. For most acute events, exposure windows of 0–1, 1–14, 1–30, and 1–42 days are often used [334,340,344]. This study design provides a direct estimate of effect (the incidence rate ratio, IRR), is well suited for rare exposures

(but not rare outcomes), and can be used to analyze multiple outcomes [390,391]. Matching on age and calendar time helps to adjust for time-varying variables that can confound the results when the vaccine and outcome are either seasonal or highly dependent on age. When the outcome is rare, however, the cohort design can be costly to implement and, for childhood vaccines that are universally recommended, there may be too few unvaccinated children for the comparison group. The design is also susceptible to selection bias that can be introduced by comparing vaccinated and unvaccinated populations, as these groups may differ by factors frequently missing from LLDBs such as ethnicity, socioeconomic status, and underlying health state [140].

In contrast to the cohort design, case–control studies are conducted by first identifying individuals who experienced a particular event over a predefined time period. This group of cases is then compared to a control group of outcome-free individuals from the same time period. Cases are often matched to controls by variables such as sex, age, study site, and calendar time [65,142,144]. This design tends to be more economical than the cohort design, and it is well suited for rare illnesses. As with the cohort method, however, the case–control design is limited when vaccine coverage rates are high and few unvaccinated cases and controls are available for analysis. In contrast to the cohort design, matching on confounding variables in a case–control study will bias the results to the null hypothesis (i.e., toward no effect) if not explicitly adjusted for in the analysis [390]. It is also difficult (and sometimes impossible) to estimate the attributable risk in a case–control design.

To address some of these limitations, alternative methods known as the risk-interval (or vaccinated cohort) and self-controlled case series (SCCS) study designs have been developed for vaccine safety epidemiology [97,146,392,393]. These designs differ from more traditional epidemiologic methods in that time intervals both before and after vaccination *within the same individual* are used to classify a person as exposed or unexposed. In the risk-interval design, incidence rates for risk and nonrisk time periods are compared, but only vaccinated individuals are included in the analysis. A time period shortly following vaccination is defined as the exposed risk interval, and events that occur during this period are classified as exposed cases. Time periods outside of the risk interval – before and after the vaccination – are considered the nonrisk (or unexposed) periods, in which occurrences of events are classified as unexposed cases. Because only vaccinated individuals are included in the study, the design eliminates biases associated with fixed factors that remain constant over time in the same individual, but differ between vaccinated and unvaccinated populations. In addition, because control time periods both before vaccination and after the risk period are included in the analysis, the design is used to examine the risk of acute, self-limiting events following vaccination.

The SCCS method is a similar design in which incidence rates for risk and nonrisk time periods are compared, but only cases with an event are included in the analysis [146,393,394]. The study population comprises cases that occur over a predefined observation period, and each case acts as its own control, thereby controlling for both measured and unmeasured confounding variables that do not vary over time (i.e., fixed confounding). With the SCCS method, multiple occurrences of independent events within an individual can be analyzed. Since only cases are required for the analysis, the SCCS study population is considerably smaller than that of the cohort, case–control, and risk-interval designs. As discussed shortly, the SCCS has nearly as much statistical power as the cohort approach when a high proportion of the population is vaccinated. Both the risk-interval and SCCS designs are analyzed with conditional Poisson regression to generate incidence rate ratios (IRRs).

Possible limitations of the risk-interval and SCCS methods stem from their inability to implicitly control for time-varying confounders, such as seasonality or age. In contrast to the matched cohort analysis, these time-varying variables must be explicitly defined as either continuous functions or categorical variables and added to parametric Poisson regression models [392,395]. Misspecifying such variables can lead to biased results – particularly when the event is rare [393]. Alternatively, it has also been shown that semi-parametric Poisson regression models can be used to analyze SCCS data in which the time-varying effects of age do not have to be explicitly defined before analysis [396].

An additional important limitation of the SCCS is that bias can be introduced if the occurrence of an event influences the probability of receiving vaccination. For example, individuals with a history of contraindicating or precautionary conditions to vaccination – such as GBS, idiopathic thrombocytopenia, anaphylaxis, and HIV – may have their immunizations either delayed or withheld indefinitely. In such a situation, the SCCS design would be limited, since only cases (i.e., those with an event) are followed forward in time, and time periods before vaccination could not be included in the analysis. This assumption of event-independent exposure (vaccination) is not required for the more traditional epidemiologic methods, because vaccination status is ascertained retrospectively from the date of diagnosis in a case–control study, and the onset of an event is ascertained prospectively from the date of vaccination in a cohort study. An analytic method has been developed to account for the post-event dependence in an SCCS analysis when the postvaccination risk period is short and when the event is both rare and nonrecurrent [397]. Simulation analyses demonstrated that the estimation method helped to correct for bias associated with event-dependent exposures, but it also produced IRR estimates that were attenuated to the null hypothesis (i.e., they underestimated the true effect). Future research is needed to develop this analytic technique further.

The characteristics of cohort, case–control, risk-interval, and SCCS designs have been compared empirically with simulation studies [392,398]. In a study using VSD data and simulated cases of a rare, acute illness (immune thrombocytopenic purpura or ITP) after MMR vaccination [340], the risk-interval, SCCS, and case–control study designs produced valid IRR estimates that were within 3% of a cohort gold standard. The case–control design, however, produced estimates that were less powerful, less precise, and biased by unmeasured fixed confounding when compared to the other study designs. The SCCS and risk-interval, in contrast, were as powerful as the cohort design and produced unbiased estimates in the presence of unmeasured fixed confounding. Of note, the SCCS design displayed similar characteristics to those of the risk-interval and cohort, but required only a fraction (0.01%) of the study population for analysis. On average, the size of the simulated cohort, risk-interval, and SCCS study populations were 2.7 million, 1.4 million, and 200 individuals, respectively.

Using similar simulation analyses, the characteristics of these four designs were evaluated in the context of real-time, active surveillance of AEFI [148]. When the exposure and outcome were acute, the cohort proved to be the best study design for active surveillance, in terms of bias, statistical power, and signal detection time. When selection bias was a concern, the risk-interval design was shown to be valid alternative. Of all the designs, the case–control design had the longest signal detection time and most biased relative risk estimates. Although the SCCS lagged behind the cohort and risk-interval designs in signal detection time, it was acceptably accurate and powerful and required only a minimum of data. Thus, the results from these simulation studies demonstrate that the SCCS

design is a valid, powerful, and economical epidemiologic tool for studying vaccine safety.

Clearly, the current methods for studying vaccine safety have contrasting strengths and limitations. In some instances, researchers employ multiple methods to address the various factors that can bias the results [98,340,389,395]. Studying the safety of the influenza vaccine, as an example, poses multiple methodologic challenges that cannot be addressed with one particular design. In a typical influenza season, more than 85% of the vaccines are administered in October and November [389,399]. It is also likely that certain conditions of interest – such as febrile seizures, gastrointestinal disorders, or rash – have a seasonal distribution across the influenza season from October through April, with the incidence peaking in winter months. Such distributions would make season a strong confounder, as it would be highly associated with both vaccination and the outcome of interest. The correlation may, in fact, be so high that one could not disentangle the individual effects of vaccination and season in the analysis. Although little can be done to rectify this potential problem with any design, the SCCS and risk-interval designs are particularly susceptible to this type of seasonal bias.

A newer strategy to account for seasonal time-varying bias in vaccine safety studies is the case-centered approach [399]. It was initially developed for assessing the effectiveness of influenza vaccines in the elderly, and has been adapted to evaluate the safety of vaccines in postmarketing settings [400]. In cases occurring during an influenza season, the method uses data from the entire cohort (cases and noncases) to calculate the probability of exposure (vaccination) for the day of the event. The logit of this probability is then placed into logistic regression model as an offset term. In essence, this method provides a seasonal adjustment for exposure by conditioning on the odds of vaccination over the course of an influenza season. This strategy has been used to study the risk of GBS following the monovalent inactivated and seasonal influenza vaccines [401] and various vaccine safety hypotheses [147,402,403].

In addition to seasonality, studying the safety of influenza vaccination is challenged by the potential for selection bias. Even where the influenza vaccine is universally recommended, it is possible that individuals who receive influenza vaccination are different from those who do not. Moreover, in large-linked MCO databases, it is possible that a certain proportion of the population received influenza vaccination outside of the MCO, which may not be captured in the automated databases [389]. As described earlier, this potential for selection bias and exposure misclassification is problematic for the cohort and case–control designs.

Impact of Larger Sample Size

The continued accumulation of data from larger study populations with common data dictionaries (e.g., by more sites and/or calendar time) has allowed new scientific evidence to emerge on longstanding vaccine safety questions.

Stratification of Risk Groups in Association Studies

Despite SIDS being reported among millions of DTP vaccinees annually since 1933, the likely association of a "temporal shift" phenomenon (Figure 20.1) was only unraveled recently with the availability of nationwide large-linked databases in Taiwan. A study assembled ~2400 SIDS cases over 10 years (~tenfold larger than most past studies), allowed stratification of risk by days since vaccination with statistically significant results for the first time (RR = 1.66; 95% confidence interval: 1.05–2.60) for the first 24 hours postvaccination, followed by a compensatory decreased RR < 1.0 for the subsequent postvaccination periods) [138]. Similarly, assembling a cohort of all recipients of measles-containing vaccinations at VSD sites over 10 years allowed analyses showing that the risk of fever and seizures was lower among vaccinees 12–15 months of age versus 16–23 months old [404].

Studies of Very Rare Outcomes/Associations

While vaccine-induced anaphylaxis and GBS were known to be rare, their true rates were difficult to ascertain. Combining LLDBs to identify potential cases that meet the Brighton Collaboration case definition for each AEFI have finally allowed such rates and association to be estimated with greater precision [57,405–407] than classical meta-analysis without Brighton validation [408,409]. For serious outcomes usually requiring hospitalization like GBS, combining hospitalization data and SCCS methods allows contribution of data from more sources [410].

The Future

Although considerable progress has been made in the development of vaccine safety analytic methods, several challenges remain. Areas of particular importance include (i) identifying optimal risk windows; (ii) characterizing a lifetime dose–response relationship from multiple influenza and tetanus-containing vaccinations; (iii) evaluating the safety of the entire childhood immunization schedule; and (iv) data mining for unknown AEFI in real-time active surveillance.

Although risk window lengths are often based on prior biologic knowledge, they are also somewhat arbitrarily defined (e.g., 0–1, 1–14, 1–42 days after vaccination). Inaccurate specification of the risk window can result either in including the true control period in the risk window or including a segment of the risk window in the control period, both of which would introduce bias. After an elevated risk has been identified in a prespecified risk window, a two-step data-driven approach to identify the period of greatest risk has been proposed. Step 1 begins by specifying a minimum risk window length, for which a risk estimate is calculated using an appropriate regression model. The risk window is incrementally lengthened and risk estimates are generated for each subsequent window. The risk estimates are plotted against the variable risk window lengths, and the researcher notes where risk is maximized. If the specified risk window is longer than the true risk window, an analytic approach is possible in step 2. Preliminary simulation and theoretical work applied to the SCCS design with conditional Poisson regression has shown that there is a linear relationship between the calculated risk and risk window length [411]. The analytic approach calculates an optimal risk window length based on maximum likelihood methods and the study design of interest. Future work should focus on applying the approach to other study designs and regression models.

Unlike all other vaccines, influenza vaccine is administered on an annual basis indefinitely. It is currently not known if the risk of certain AEFI increases with each subsequent dose. For children in particular, studying this relationship is problematic, since dose number is likely to be strongly correlated with age. In other words, since both age at vaccination and cumulative dose increase over time, it would be difficult to explain how much of the risk associated with a particular dose can be explained by age. To study adequately the relationship between dose number and the risk of AEFI, new methods for disentangling the correlation between dose and age are needed.

An increasing number of parents are choosing to either delay or refuse some or all vaccines for their children, often citing safety concerns [412–415]. These sentiments reflect the number, frequency, and timing of vaccines, leading parents to undervaccinate and adopt immunization schedules that differ from the ACIP recommended schedule, known as the alternative immunization schedule [141,415–419]. In its 2013 report, IOM/NAM recommended that additional observational research was warranted to compare the health outcomes between children vaccinated according to the

ACIP schedule and those on an alternative schedule [420]. Using observational studies for these comparisons could be methodologically challenging, because health- and healthcare–seeking behaviors may differ systematically between the two groups [141,421]. In addition, it involves defining exposure to multiple different immunization schedules, identifying biologically plausible AEFI outcomes to study in the context of the immunization schedule, and identifying epidemiologic and statistical methods to assess the safety of the schedule as a whole [133].

Defining exposure is particularly challenging, because even though undervaccination is common, there is considerable variability in patterns of undervaccination and there could be multiple ways to assess it. A 2013 study showed that 48.7% of children under age 2 years were undervaccinated for at least one day, and there were 1399 different patterns of undervaccination using the eight recommended childhood vaccines and categorizing them into three groups of all doses received on-time, no doses received, or some doses either missing or not received on time [141]. These patterns did not consider other factors related to the schedule such as the age, spacing, or order of vaccinations, which would result in millions of different combinations of vaccination patterns. Different approaches for identifying patterns of undervaccination have been proposed using the average days undervaccinated and proportion of days undervaccinated metrics, which quantify the number of calendar days for which a child is undervaccinated across all recommended vaccines [141,422–424]. Another problem with defining the exposure is the potential for misclassification of the vaccination status in EHRs. This could be minimized if there is an ICD code for vaccine refusal in the EHR, but these children would likely be too few to be able to study rare AEFI [425].

In the context of the entire immunization schedule, the outcomes to be studied need to be based on biologic plausibility, feasibility in the context of the overall immunization schedule, epidemiologic evidence, public health significance, and public concern [133,420]. Studying short-term acute outcomes poses challenges, as they may occur before the completion of the schedule, and are typically associated with short risk periods following specific vaccines, doses, or combination of vaccines. Moreover, short-term outcomes may influence parents' future decision to vaccinate their child and may lead to reverse causality.

The VSD vaccine schedule study team along with subject matter experts have prioritized a list of 20 plausible long-term or chronic outcomes [133]. Some of the outcomes on this list are very rare events, such as meningitis, encephalopathy, and development of type 1 diabetes, and very large cohorts of children will have to be studied to be able to identify differences in risk. More common outcomes on the list, such as asthma and allergy development, would entail intensive primary data collection resources to address the potential for misclassification. A particular challenge is the study of outcomes that have an insidious onset. Early symptoms of disease may alter parents' vaccination decision for their child, but before any formal clinical diagnosis has been made, leading to reverse causality [133].

Deciding on study designs and analytic plans to evaluate the safety of the schedule can be complex, because these may be susceptible to selection bias, confounding, and misclassification. Undervaccinated children differ from children who receive the recommended vaccines on time on important variables such as baseline health and healthcare utilization, race/ethnicity, socioeconomic status, parental education, and family history of illness. These data are difficult to capture using EHRs and should be addressed using methods including directed acyclic graphs (DAGs), control outcomes, restriction, matching, primary data collection, and sensitivity analyses [133].

Lastly, data-mining methods have been developed to identify signals for unexpected AEFI. These methods for vaccine safety have been

applied to passive surveillance systems and *ad hoc* epidemiologic studies [344,426]. In these respective settings, however, data-mining analyses have been limited by reporting bias, lack of denominator data, and low statistical power for rare events. Conducting data-mining analyses in LLDBs with real-time active surveillance will address some of these limitations. Such methods would be a natural complement to targeted active surveillance, in which AEFI are specified *a priori*. For targeted active surveillance (see Chapter 46), sequential testing methods have been developed to protect against false positive signals (type I error) when data are analyzed on a weekly or monthly basis. The potential for type I error, however, will increase significantly when multiple unspecified outcomes are analyzed at the same time.

New analytic tools for using LLDBs to identify unsuspected AEFI in real time have recently been developed [360,427]. These methods should be sensitive enough to detect potentially serious AEFI, but also conservative enough to protect against too many false signals. Such methods also need to account for seasonality, selection biases, and other factors that can distort the findings. Perhaps most importantly, a process for signal validation (e.g., controlled epidemiologic studies with medical chart review) and a plan for risk communication must be in place should a signal arise [428].

More broadly, major integration (and possible reorganization) of the vaccinology enterprise writ large may be needed to allow a shift from a historically largely empiric "one size fits all" paradigm to one allowing "personalized medicine" requiring knowledge in various "systems biology" processes [117]. One such process, adversomics, hopes to identify, characterize, and predict adverse or maladaptive immune responses [125]. To do so will eventually require linkage of basic science, clinical trial, and postmarketing surveillance databases, with appropriate analytic tools and graphic displays. Gaps in knowledge or processes identified (e.g., biobanking of samples from persons with serious AEFI) will need resolution. As ongoing immunizations of a substantial proportion of the human population against an ever-increasing list of VPDs globally is probable, both the sample size and logistic requirements are manageable, so the challenges are likely to be more financial and political. Advances in genomics may allow other improvements in vaccine safety. Deep sequencing assays can detect DNA/RNA sequences that have contaminated the vaccine production process [429], help document exposure to them in vaccinees, or monitor for circulation of risky live vaccine strains [430]. Homology between vaccine candidate and human genomic sequences can identify potential causes of autoimmunity [431].

As recently as 2012, only one-third of WHO member states had an effective vaccine safety monitoring system [206]. While progress has been made, significant challenges in capacity building remain – to initially establish an SRS, followed by effective response to any signals detected [209]. The recent availability of funding from the Global Alliance for Vaccines and Immunizations (GAVI) for these purposes in eligible LMICs should hopefully turn the tide [432].

References

1 Feikin DR, Flannery B, Hamel MJ, Stack M, Hansen PM. Vaccines for children in low- and middle-income countries. In Black RE, Laxminarayan R, Temmerman M, Walker N, eds., *Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities*, vol. 2. 3rd edn. Washington, DC: International Bank for Reconstruction and Development/World Bank, 2016, pp. 187–204.

2 Ozawa S, Clark S, Portnoy A, *et al*. Estimated economic impact of vaccinations in 73 low- and middle-income countries, 2001–2020. *Bull World Health Organ* 2017; **95**(9): 629–38.

3 Prevention CfDCa. Achievements in public health, 1900–1999 impact of vaccines universally recommended for children – United States, 1990–1998. *Morb Mortal Wkly Rep* 1999; **48**(12): 243–8.

4 WHO, UNICEF, World Bank. *State of the World's Vaccines and Immunization*, 3rd edn. Geneva: World Health Organization, 2009.

5 Wilson GS. *The Hazards of Immunization*. London: Athlone Press, 1967.

6 Chen RT. Vaccine risks: real, perceived and unknown. *Vaccine* 1999; **17**(Suppl 3): S41–S46.

7 Plotkin SA. Lessons learned concerning vaccine safety. *Vaccine* 2001; **20**(suppl 1): S16–S19.

8 Duclos P, Delo A, Aguado T, *et al*. Immunization safety priority project at the World Health Organization. *Semin Pediatr Infect Dis* 2003; **14**(3): 233–9.

9 Cooper LZ, Larson HJ, Katz SL. Protecting public trust in immunization. *Pediatrics* 2008; **122**(1): 149–53.

10 Asturias EJ, Wharton M, Pless R, *et al*. Contributions and challenges for worldwide vaccine safety: the Global Advisory Committee on Vaccine Safety at 15 years. *Vaccine* 2016; **34**(29): 3342–9.

11 Dubé E, Gagnon D, MacDonald N, *et al*. Underlying factors impacting vaccine hesitancy in high income countries: a review of qualitative studies. *Expert Rev Vaccines* 2018; **17**(11): 989–1004.

12 Gangarosa EJ, Galazka AM, Wolfe CR, *et al*. Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet* 1998; **351**(9099): 356–61.

13 Orenstein WA. DTP vaccine litigation, 1988. *Am J Dis Child* 1990; **144**(5): 517.

14 Institute of Medicine. Liability for the production and sale of vaccine. In Sanford JP, ed. *Vaccine Supply and Innovation*. Washington, DC: National Academies Press, 1985, pp. 85–122.

15 Ball LK, Ball R, Pratt RD. An assessment of thimerosal use in childhood vaccines. *Pediatrics* 2001; **107**(5): 1147–54.

16 Verstraeten T, Davis RL, DeStefano F, *et al*. Safety of thimerosal-containing vaccines: a two-phased study of computerized health maintenance organization databases. *Pediatrics* 2003; **112**(5): 1039–48.

17 Joellenbeck LM, Zwanziger LL, Durch JS, Strom BL. *The Anthrax Vaccine: Is It Safe? Does It Work?* Washington, DC: National Academies Press, 2002.

18 Poland GA, Grabenstein JD, Neff JM. The US smallpox vaccination program: a review of a large modern era smallpox vaccination implementation program. *Vaccine* 2005; **23**(17–18): 2078–81.

19 Wakefield AJ, Murch SH, Anthony A, *et al*. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 1998; **351**(9103): 637–41.

20 Mahase E. Measles cases rise 300% globally in first few months of 2019. *BMJ* 2019; **365**: 1810.

21 Balinska MA. Hepatitis B vaccination and French society ten years after the suspension of the vaccination campaign: how should we raise infant immunization coverage rates? *J Clin Virol* 2009; **46**(3): 202–5.

22 Clements CJ, Greenough P, Shull D. How vaccine safety can become political: the example of polio in Nigeria. *Curr Drug Saf* 2006; **1**(1): 117–19.

23 Ghinai I, Willott C, Dadari I, Larson HJ. Listening to the rumours: what the northern Nigeria polio vaccine boycott can tell us ten years on. *Glob Public Health* 2013; **8**(10): 1138–50.

24 Chandler RE. Safety concerns with HPV vaccines continue to linger: are current vaccine pharmacovigilance practices sufficient? *Drug Saf* 2017; **40**(12): 1167–70.

25 Schwarzinger M, Flicoteaux R, Cortarenoda S, Obadia Y, Moatti JP. Low acceptability of A/H1N1 pandemic vaccination in French

adult population: did public health policy fuel public dissonance? *PLoS One* 2010; **5**(4): e10199.

26 Eastwood K, Durrheim DN, Jones A, Butler M. Acceptance of pandemic (H1N1) 2009 influenza vaccination by the Australian public. *Med J Aust* 2010; **192**(1): 33–6.

27 Hilton S, Petticrew M, Hunt K. "Combined vaccines are like a sudden onslaught to the body's immune system": parental concerns about vaccine "overload" and "immune-vulnerability". *Vaccine* 2006; **24**(20): 4321–7.

28 Kennedy A, Lavail K, Nowak G, Basket M, Landry S. Confidence about vaccines in the United States: understanding parents' perceptions. *Health Aff (Millwood)*. 2011; **30**(6): 1151–9.

29 Lieu TA, Zikmund-Fisher BJ, Chou C, Ray GT, Wittenberg E. Parents' perspectives on how to improve the childhood vaccination process. *Clin Pediatr (Phila)* 2017; **56**(3): 238–46.

30 Fineberg HV, Howe CJ, Howson CP. *Adverse Effects of Pertussis and Rubella Vaccines*. Washington, DC: National Academies Press, 1991.

31 Stratton KR, Howe CJ, Johnston RB. *Adverse Events Associated with Childhood Vaccines: Evidence Bearing on Causality*. Washington, DC: National Academies Press, 1994.

32 Immunization Safety Review Committee. *Immunization Safety Review: Thimerosal-Containing Vaccines and Neurodevelopmental Disorders*. Stratton K, Gable A, McCormick M, eds. Washington, DC: National Academies Press, 2001.

33 Immunization Safety Review Committee. *Immunization Safety Review: Measles–Mumps–Rubella Vaccine and Autism*. Stratton K, Gable A, Shetty P, McCormick M, eds. Washington, DC: National Academies Press, 2001.

34 Immunization Safety Review Committee. *Immunization Safety Review: Multiple Immunizations and Immune Dysfunction*. Stratton K, Wilson CB, McCormick M, eds. Washington, DC: National Academies Press, 2002.

35 Immunization Safety Review Committee. *Immunization Safety Review: SV40 Contamination of Polio Vaccine and Cancer*. Stratton K, McCormick M, eds. Washington, DC: National Academies Press, 2002.

36 Immunization Safety Review Committee. *Immunization Safety Review: Hepatitis B Vaccine and Demyelinating Neurological Disorders*. Stratton K, Almario D, McCormick M, eds. Washington, DC: National Academies Press, 2002.

37 Immunization Safety Review Committee. *Immunization Safety Review: Vaccinations and Sudden Unexpected Death in Infancy*. Stratton K, Almario DA, Wizeman TM, McCormick MC, eds. Washington, DC: National Academies Press, 2003.

38 Immunization Safety Review Committee. *Immunization Safety Review: Influenza Vaccines and Neurological Complications*. Stratton K, Almario DA, Wizeman TM, McCormick MC, eds. Washington, DC: National Academies Press, 2003.

39 Immunization Safety Review Committee. *Immunization Safety Review: Adverse Effects of Vaccines: Evidence and Causality*. Stratton K, Ford A, Rusch E, Clayton EW, eds. Washington, DC: National Academies Press, 2011.

40 Hinshaw A, Aragon T, Berg A, *et al*. *Childhood Immunization Schedule and Safety: Stakeholder Concerns, Scientific Evidence, and Future Studies*. Washington, DC: National Academies Press, 2013. http://www.nationalacademies.org/hmd/Reports/2013/The-Childhood-Immunization-Schedule-and-Safety.aspx (accessed May 2019).

41 Griffin MR, Braun MM, Bart KJ. What should an ideal vaccine postlicensure safety system be? *Am J Public Health* 2009; **99**(Suppl 2): S345–S350.

42 Lopalco PL, Johansen K, Ciancio B, De Carvalho GH, Kramarz P, Giesecke J. Monitoring and assessing vaccine safety: a European perspective. *Expert Rev Vaccines* 2010; **9**(4): 371–80.

43 Bonhoeffer J, Black S, Izurieta H, Zuber P, Sturkenboom M. Current status and future directions of post-marketing vaccine safety monitoring with focus on USA and Europe. *Biologicals* 2012; **40**(5): 393–7.

44 Chen RT, Shimabukuro TT, Martin DB, Zuber PL, Weibel DM, Sturkenboom M. Enhancing vaccine safety capacity globally: a lifecycle perspective. *Am J Prev Med* 2015; **49**(6 Suppl 4): S364–S376.

45 Gruber MF, Marshall VB. Regulation and testing of vaccines. In Plotkin S, Orenstein WA, Offit PA, Edwards KM, eds. *Plotkin's Vaccines*, 7th edn. Philadelphia, PA: Elsevier, 2017, pp. 1547–65.

46 Ahmed SS, Ellis RW, Rappuoli R,.Technologies for making new vaccines. In Plotkin S, Orenstein WA, Offit PA, Edwards KM, eds. *Plotkin's Vaccines*, 7th edn. Philadelphia, PA: Elsevier, 2017, pp. 1283–304.

47 Jackson LA, Yu O, Nelson J, *et al*. Risk of medically attended local reactions following diphtheria toxoid containing vaccines in adolescents and young adults: a Vaccine Safety Datalink study. *Vaccine* 2009; **27**(36): 4912–16.

48 Bernstein DI, Edwards KM, Dekker CL, *et al*. Effects of adjuvants on the safety and immunogenicity of an avian influenza H5N1 vaccine in adults. *J Infect Dis* 2008; **197**(5): 667–75.

49 Lewis DJ, Huo Z, Barnett S, *et al*. Transient facial nerve paralysis (Bell's palsy) following intranasal delivery of a genetically detoxified mutant of Escherichia coli heat labile toxin. *PLoS One* 2009; **4**(9): e6999.

50 Zent O, Hennig R. Post-marketing surveillance of immediate allergic reactions: polygeline-based versus polygeline-free pediatric TBE vaccine. *Vaccine* 2004; **23**(5): 579–84.

51 Novadzki IM, Rosario N, Zanoni G, Chiesa E, Puccetti A, Simone R. Hypersensitivity reactions following measles–mumps–rubella vaccine and dextran-specific IgG response. *Allergol Immunopathol (Madr)* 2010; **38**(6): 341–3.

52 Parthasarathy A. Contamination of reconstituted multidose measles vaccine vial and toxic shock syndrome in Tamilnadu. *Indian Pediatr* 2008; **45**(7): 606.

53 Offit PA, Jew RK. Addressing parents' concerns: do vaccines contain harmful preservatives, adjuvants, additives, or residuals? *Pediatrics* 2003; **112**(6 Pt 1): 1394–7.

54 Pepin J, Abou Chakra CN, Pepin E, Nault V, Valiquette L. Evolution of the global burden of viral infections from unsafe medical injections, 2000–2010. *PLoS One* 2014; **9**(6): e99677.

55 Cousins S. Contaminated vaccine deaths a serious setback for Syria. *Lancet* 2014; **384**(9949): 1172.

56 Schonberger LB, Bregman DJ, Sullivan-Bolyai JZ, *et al*. Guillain-Barré syndrome following vaccination in the National Influenza Immunization Program, United States, 1976–1977. *Am J Epidemiol* 1979; **110**(2): 105–23.

57 Salmon DA, Proschan M, Forshee R, *et al.* Association between Guillain-Barré syndrome and influenza A (H1N1) 2009 monovalent inactivated vaccines in the USA: a meta-analysis. *Lancet* 2013; **381**(9876): 1461–8.

58 Platt LR, Estivariz CF, Sutter RW. Vaccine-associated paralytic poliomyelitis: a review of the epidemiology and estimation of the global burden. *J Infect Dis* 2014; **210**(Suppl 1): S380–S389.

59 McGoldrick MD, Bailie GR. Nonnarcotic analgesics: prevalence and estimated economic impact of toxicities. *Ann Pharmacother* 1997; **31**(2): 221–7.

60 Marcuse EK, Wentz KR. The NCES reconsidered: summary of a 1989 workshop. National Childhood Encephalopathy Study. *Vaccine* 1990; **8**(6): 531–5.

61 Miller D, Madge N, Diamond J, *et al*. Pertussis immunisation and serious acute neurological illnesses in children. *BMJ* 1993; **307**: 1171–6.

62 Wentz KR, Marcuse EK. Diphtheria–tetanus–pertussis vaccine and serious neurologic illness: an updated review of the epidemiologic evidence. *Pediatrics* 1991; **87**(3): 287–97.

63 Scheffer IE. Vaccination triggers, rather than causes, seizures. *Epilepsy Curr* 2015; **15**(6): 335–7.

64 Safranek TJ, Lawrence DN, Kurland LT, *et al*. Reassessment of the association between Guillain-Barré syndrome and receipt of swine influenza vaccine in 1976–1977: results of a two-state study. Expert Neurology Group. *Am J Epidemiol* 1991; **133**(9): 940–51.

65 Murphy TV, Gargiullo PM, Massoudi MS, *et al*. Intussusception among infants given an oral rotavirus vaccine. *N Engl J Med* 2001; **344**(8): 564–72.

66 Kramarz P, France EK, DeStefano F, *et al*. Population-based study of rotavirus vaccination and intussusception. *Pediatr Infect Dis J* 2001; **20**(4): 410–16.

67 Simonsen L, Morens D, Elixhauser A, Gerber M, Van Raden M, Blackwelder W. Effect of rotavirus vaccination programme on trends in admission of infants to hospital for intussusception. *Lancet* 2001; **358**(9289): 1224–9.

68 Rha B, Tate JE, Weintraub E, *et al*. Intussusception following rotavirus vaccination: an updated review of the available evidence. *Expert Rev Vaccines* 2014; **13**(11): 1339–48.

69 McNeil MM, Gee J, Weintraub ES, *et al*. The Vaccine Safety Datalink: successes and challenges monitoring vaccine safety. *Vaccine* 2014; **32**(42): 5390–8.

70 Hviid A. Postlicensure epidemiology of childhood vaccination: the Danish experience. *Expert Rev Vaccines* 2006; **5**(5): 641–9.

71 Levine EM, Davey AS, Houston AM. Legal issues. In Plotkin S, Orenstein WA, Offit PA, Edwards KM, eds. *Plotkin's Vaccines*, 5th edn. Philadelphia, PA: Elsevier, 2017, pp. 1601–30.

72 John TJ, Samuel R. Herd immunity and herd effect: new insights and definitions. *Eur J Epidemiol* 2000; **16**(7): 601–6.

73 Hardin G. The tragedy of the commons. *Science* 1968; **162**(5364): 1243–8.

74 Fine PE, Clarkson JA. Individual versus public priorities in the determination of optimal vaccination policies. *Am J Epidemiol* 1986; **124**(6): 1012–20.

75 Manfredi P, Posta PD, d'Onofrio A, *et al*. Optimal vaccination choice, vaccination games, and rational exemption: an appraisal. *Vaccine* 2009; **28**(1): 98–109.

76 Patel M, Lee AD, Redd SB, *et al*. Increase in measles cases - United States, January 1-April 26, 2019. *MMWR Morb Mortal Wkly Rep* 2019; **68**(17): 402–4. doi: 10.15585/mmwr.mm6817e1

77 Nathanson N, Langmuir AD. The Cutter incident. Poliomyelitis following formaldehyde-inactivated poliovirus vaccination in the United States during the Spring of 1955. II. Relationship of poliomyelitis to Cutter vaccine. 1963. *Am J Epidemiol* 1995; **142**(2): 109–40.

78 Klug B, Robertson JS, Condit RC, *et al*. Adventitious agents and live viral vectored vaccines: considerations for archiving samples of biological materials for retrospective analysis. *Vaccine* 2016; **34**(51): 6617–25.

79 Food and Drug Administration. Bovine derived materials used in vaccine manufacturing: questions and answers. 2013. https://www.fda.gov/biologicsbloodvaccines/vaccines/questionsaboutvaccines/ucm143521.htm (accessed May 2019).

80 Victoria JG, Wang C, Jones MS, *et al*. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *J Virol* 2010; **84**(12): 6033–40.

81 Ehmann F, Kurz X, Cavaleri M, Arlett PR. Regulation of vaccines in Europe. In Plotkin SA, Orenstein WA, Offit PA, Edwards KM, eds. *Plotkin's Vaccines*, 7th edn. Philadelphia, PA: Elsevier, 2017, pp. 1566–72.

82 Belgharbi L, Broojerdi AK, Rodriguez-Hernandez CA, Wood D. Regulation of vaccines in low- and middle-income countries. In Plotkin SA, Orenstein WA, Offit PA, Edwards KM, eds. *Plotkin's Vaccines*, 7th edn. Philadelphia, PA: Elsevier, 2017, pp. 1573–83.

83 Khan AS, Vacante DA, Cassart JP, *et al*. Advanced Virus Detection Technologies Interest Group (AVDTIG): efforts on high throughput sequencing (HTS) for virus detection. *PDA J Pharm Sci Technol* 2016; **70**(6): 591–5.

84 Neustadt RE, Fineberg HV. *The Swine Flu Affair: Decision-Making on a Slippery Disease*. Washington, DC: US Government Printing Office, 1978.

85 Prevention CfDCa. Poliomyelitis prevention in the United States: updated recommendations of the Advisory Committee on Immunization Practices (ACIP). *Morb Mortal Wkly Rep*, 2000; **49**(RR–5): 1–25.

86 Two MMR vaccines withdrawn. *Lancet* 1992; **340**: 922.

87 Colville A, Pugh S, Miller E. Withdrawal of a mumps vaccine. *Eur J Pediatr* 1994; **153**(6): 467–8.

88 Lathrop SL, Ball R, Haber P, *et al*. Adverse event reports following vaccination for Lyme disease: December 1998–July 2000. *Vaccine* 2002; **20**(11–12): 1603–8.

89 Andreae MC, Freed GL, Katz SL. Safety concerns regarding combination vaccines: the experience in Japan. *Vaccine* 2004; **22**(29–30): 3911–16.

90 Halabi SF, Omer SB. A global vaccine injury compensation system. *JAMA* 2017; **317**(5): 471–2.

91 Human Resources and Services Administration. National vaccine injury compensation program: revisions to the vaccine injury table. Final rule. *Fed Regist* 2017; **82**(12): 6294–305.

92 Lennon D, Jackson C, Wong S, Horsfall M, Stewart J, Reid S. Fast tracking the vaccine licensure process to control an epidemic of serogroup B meningococcal disease in New Zealand. *Clin Infect Dis* 2009; **49**(4): 597–605.

93 Centers for Disease Control and Prevention. Public Health Service recommendations on smallpox vaccination. *Morb Mortal Wkly Rep* 1971; **20**: 339–45.

94 Lopalco PL. Wild and vaccine-derived poliovirus circulation, and implications for polio eradication. *Epidemiol Infect* 2017; **145**(3): 413–19.

95 Peltola H, Kulkarni PS, Kapre SV, Paunio M, Jadhav SS, Dhere RM. Mumps outbreaks in Canada and the United States: time for new thinking on mumps vaccines. *Clin Infect Dis* 2007; **45**(4): 459–66.

96 Henderson DA, Arita I. The smallpox threat: a time to reconsider global policy. *Biosecur Bioterror* 2014; **12**(3): 117–21.

97 Farrington CP, Pugh S, Colville A, *et al*. A new method for active surveillance of adverse events from diphtheria/tetanus/pertussis and measles/mumps/rubella vaccines. *Lancet* 1995; **345**(8949): 567–9.

98 Mutsch M, Zhou W, Rhodes P, *et al*. Use of the inactivated intranasal influenza vaccine and the risk of Bell;s palsy in Switzerland. *N Engl J Med* 2004; **350**(9): 896–903.

99 Jacobsen SJ, Ackerson BK, Sy LS, *et al*. Observational safety study of febrile convulsion following first dose MMRV vaccination in a managed care setting. *Vaccine* 2009; **27**(34): 4656–61.

100 Stetler HC, Orenstein WA, Bart KJ, Brink EW, Brennan JP, Hinman AR. History of convulsions and use of pertussis vaccine. *J Pediatr* 1985; **107**(2): 175–9.

101 Martin M, Weld LH, Tsai TF, *et al*. Advanced age a risk factor for illness temporally associated with yellow fever vaccination. *Emerg Infect Dis* 2001; **7**(6): 945–51.

102 Oliveira AC, Maria Henrique da Mota L, Dos Santos-Neto LL, *et al*. Occurrence of autoimmune diseases related to the vaccine against yellow fever. *Autoimmune Dis* 2014; **2014**: 473170.

103 Pichichero ME. Acellular pertussis vaccines: towards an improved safety profile. *Drug Saf* 1996; **15**(5): 311–24.

104 Francis T, Korns R, Voight R. An evaluation of the 1954 poliomyelitis vaccine trials: summary report. *Am J Public Health* 1955; **45**(5 Part 2): 1–63.

105 Fine PE. Methodological issues in the evaluation and monitoring of vaccine safety. *Ann N Y Acad Sci* 1995; **754**: 300–8.

106 Dos Santos BA, Ranieri TS, Bercini M, *et al*. An evaluation of the adverse reaction potential of three measles–mumps–rubella combination vaccines. *Rev Panam Salud Publica* 2002; **12**(4): 240–6.

107 Bonhoeffer J, Kohl K, Chen R, *et al*. The Brighton Collaboration: addressing the need for standardized case definitions of adverse events following immunization (AEFI). *Vaccine* 2002; **21**(3–4): 298–302.

108 Heijbel H, Ciofi degli Atti MC, Harzer E, *et al*. Hypotonic hyporesponsive episodes in eight pertussis vaccine studies. *Dev Biol Stand* 1997; **89**: 101–3.

109 Aaby P, Knudsen K, Whittle H, *et al*. Long-term survival after Edmonston-Zagreb measles vaccination in Guinea-Bissau: increased female mortality rate. *J Pediatr* 1993; **122**(6): 904–8.

110 Aaby P, Jensen H, Simondon F, Whittle H. High-titer measles vaccination before 9 months of age and increased female mortality: do we have an explanation? *Semin Pediatr Infect Dis* 2003; **14**(3): 220–32.

111 Shann F, Nohynek H, Scott JA, Hesseling A, Flanagan KL. Randomized trials to study the nonspecific effects of vaccines in children in low-income countries. *Pediatr Infect Dis J* 2010; **29**(5): 457–61.

112 Hall AJ, Cutts FT. Lessons from measles vaccination in developing countries. *BMJ* 1993; **307**(6915): 1294–5.

113 Cutts FT, Fine PE. Caution: mortality ratios ahead. *Lancet* 2003; **361**(9376): 2169–70.

114 Dans AL, Dans LF, Lansang MAD, Silvestre MAA, Guyatt GH. Review of a licensed dengue vaccine: inappropriate subgroup analyses and selective reporting may cause harm in mass vaccination programs. *J Clin Epidemiol* 2018; **95**: 137–9.

115 Izurieta HS, Zuber P, Bonhoeffer J, *et al*. Roadmap for the international collaborative epidemiologic monitoring of safety and effectiveness of new high priority vaccines. *Vaccine* 2013; **31**(35): 3623–7.

116 Black S, Mulholland K, Halsey NA. Developing sustainable funding for vaccine safety infrastructure. *Vaccine* 2010; **28**(5): 1133–4.

117 Poland GA, Ovsyannikova IG, Kennedy RB. Personalized vaccinology: a review. *Vaccine* 2018; **36**: 5350–7.

118 Hagan T, Nakaya HI, Subramaniam S, Pulendran B. Systems vaccinology: enabling rational vaccine design with systems biological approaches. *Vaccine* 2015; **33**(40): 5294–301.

119 Zanoni G, Nguyen TM, Valsecchi M, Gallo G, Tridente G. Prevention and monitoring of adverse events following immunization: the "Green Channel" of the Veneto region in Italy. *Vaccine* 2003; **22**(2): 194–201.

120 Crawford NW, Hodgson K, Gold M, Buttery J, Wood N, AEFI-CAN Network. Adverse events following HPV immunization in Australia: establishment of a clinical network. *Hum Vaccin Immunother* 2016; **12**(10): 2662–5.

121 LaRussa PS, Edwards KM, Dekker CL, *et al*. Understanding the role of human variation in vaccine adverse events: the Clinical Immunization Safety Assessment Network. *Pediatrics* 2011; **127**(Suppl 1): S65–S73.

122 Li R, McNeil MM, Pickering S, *et al*. Military healthcare providers reporting of adverse events following immunizations to the vaccine adverse event reporting system. *Mil Med* 2014; **179**(4): 435–41.

123 Shaw J, Halsey NA, Weinberg A, *et al*. Arm paralysis after routine childhood

vaccinations: application of advanced molecular methods to the causality assessment of an adverse event after immunization. *J Pediatric Infect Dis Soc* 2017; **6**(3): e161–e164.

124 Engler RJ, Nelson MR, Collins LC, Jr., *et al*. A prospective study of the incidence of myocarditis/pericarditis and new onset cardiac symptoms following smallpox and influenza vaccination. *PLoS One* 2015; **10**(3): e0118283.

125 Whitaker JA, Ovsyannikova IG, Poland GA. Adversomics: a new paradigm for vaccine safety and design. *Expert Rev Vaccines* 2015; **14**(7): 935–47.

126 Wise RP, Kiminyo KP, Salive ME. Hair loss after routine immunizations. *JAMA* 1997; **278**(14): 1176–8.

127 Centers for Disease Control and Prevention. Intussusception among recipients of rotavirus vaccine: United States, 1998–1999. *Morb Mortal Wkly Rep* 1999; **48**(27): 577–81.

128 Chen RT. Surveillance for vaccine adverse events and public health disease: similarities and differences. *Pharmacoepidemiol Drug Saf* 1996; **5**(S1): S45.

129 Division of Similarities and Immunization. Vaccine-associated adverse events in Canada, 1992 Report. *Can Commun Dis Rep* 1995; **21**: 117–28.

130 Chen RT, Rastogi SC, Mullen JR, *et al*. The Vaccine Adverse Event Reporting System (VAERS). *Vaccine* 1994; **12**(6): 542–50.

131 Miller E, Goldacre M, Pugh S, *et al*. Risk of aseptic meningitis after measles, mumps, and rubella vaccine in UK children. *Lancet* 1993; **341**(8851): 979–82.

132 Grabenstein JD. The Vaccine Identification Standards Initiative (VISI): towards clearer labels and common nomenclature. *Hospital Pharmacy* 2002; **37**(1): 58–74.

133 Glanz JM, Newcomer SR, Jackson ML, *et al*. White Paper on studying the safety of the childhood immunization schedule in the Vaccine Safety Datalink. *Vaccine* 2016; **34**(Suppl 1): A1–A29.

134 Velentgas P, Bohn RL, Brown JS, *et al*. A distributed research network model for post-marketing safety studies: the Meningococcal Vaccine Study. *Pharmacoepidemiol Drug Saf* 2008; **17**(12): 1226–34.

135 Lasky T, Terracciano GJ, Magder L, *et al*. The Guillain-Barré syndrome and the 1992–1993 and 1993–1994 influenza vaccines. *N Engl J Med* 1998; **339**(25): 1797–802.

136 Report of the working group on the possible relationship between hepatitisB vaccination and the chronic fatigue syndrome. *Can Med Assoc J* 1993; **149**: 314–19.

137 Sarkanen TO, Alakuijala APE, Dauvilliers YA, Partinen MM. Incidence of narcolepsy after H1N1 influenza and vaccinations: systematic review and meta-analysis. *Sleep Med Rev* 2018; **38**: 177–86.

138 Huang WT, Chen RT, Hsu YC, Glasser JW, Rhodes PH. Vaccination and unexplained sudden death risk in Taiwanese infants. *Pharmacoepidemiol Drug Saf* 2017; **26**(1): 17–25.

139 Thompson NP, Montgomery SM, Pounder RE, Wakefield AJ. Is measles vaccination a risk factor for inflammatory bowel disease? *Lancet* 1995; **345**(8957): 1071–4.

140 Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1992; **136**(2): 121–35.

141 Glanz JM, Newcomer SR, Narwaney KJ, *et al*. A population-based cohort study of undervaccination in 8 managed care organizations across the United States. *JAMA Pediatr* 2013; **167**(3): 274–81.

142 DeStefano F, Mullooly JP, Okoro CA, *et al*. Childhood vaccinations, vaccination timing, and risk of type 1 diabetes mellitus. *Pediatrics* 2001; **108**(6): E112.

143 Ascherio A, Zhang SM, Hernan MA, *et al*. Hepatitis B vaccination and the risk of

multiple sclerosis. *N Engl J Med* 2001; **344**(5): 327–32.

144 Black C, Kaye JA, Jick H. MMR vaccine and idiopathic thrombocytopaenic purpura. *Br J Clin Pharmacol* 2003; **55**(1): 107–11.

145 Mortimer EA, Jones P, Adelson L. DTP and SIDS – Letter. *Pediatr Infect Dis* 1983; **2**: 492–3.

146 Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995; **51**(1): 228–35.

147 Baxter R, Lewis E, Goddard K, *et al*. Acute demyelinating events following vaccines: a case-centered analysis. *Clin Infect Dis* 2016; **63**(11): 1456–62.

148 McClure DL, Glanz JM, Xu S, Hambidge SJ, Mullooly JP, Baggs J. Comparison of epidemiologic methods for active surveillance of vaccine safety. *Vaccine* 2008; **26**: 3341–5.

149 Andrews N. Epidemiological designs for vaccine safety assessment: methods and pitfalls. *Biologicals* 2012; **40**(5): 389–92.

150 Baker MA, Lieu TA, Li L, *et al*. A vaccine study design selection framework for the postlicensure rapid immunization safety monitoring program. *Am J Epidemiol* 2015; **181**(8): 608–18.

151 US Food and Drugs Administration. Toxicity grading scale for healthy adult and adolescent volunteers enrolled in preventive vaccine clinical trials. Rockville, MD: Botswana Government Print, 2007.

152 Bonhoeffer J, Bentsi-Enchill A, Chen RT, *et al*. Guidelines for collection, analysis and presentation of vaccine safety data in pre- and post-licensure clinical studies. *Vaccine* 2009; **27**(16): 2282–8.

153 Rennels MB. The rotavirus vaccine story: a clinical investigator's view. *Pediatrics* 2000; **106**(1 Pt 1): 123–5.

154 Aagaard L, Soendergaard B, Stenver DI, Hansen EH. Knowledge creation about ADRs: turning the perspective from the rear mirror to the projector? *Br J Clin Pharmacol* 2008; **65**(3): 364–76.

155 Ray WA, Griffin MR. Re: "Confounding in studies of adverse reactions to vaccines." *Am J Epidemiol* 1994; **139**(2): 229–30.

156 Ellenberg SS. Safety considerations for new vaccine development. *Pharmacoepidemiol Drug Saf* 2001; **10**(5): 411–15.

157 Lesko SM, Mitchell AA. The safety of acetaminophen and ibuprofen among children younger than two years old. *Pediatrics* 1999; **104**(4): e39.

158 Scott TF, Bonanno DE. Reactions to live-measles-virus vaccine in children previously inoculated with killed-virus vaccine. *N Engl J Med* 1967; **277**(5): 248–50.

159 VII CWG. The Development Safety Update Report (DSUR): Harmonizing the format and content for periodic safety reporting during clinical trials: Report of the CIOMS Working Group VII. Geneva: CIOMS, 2006.

160 Rock EP, Molloy VJ, Humphrey JS. GCP data quality for early clinical development. *Clin Cancer Res* 2010; **16**(6): 1756–63.

161 Mgone CS, Salami W. EDCTP: a genuine north–south partnership. *Trop Med Int Health* 2009; **14**(11): 1327–8.

162 Izurieta HS, Moro PL, Chen RT. Hospital-based collaboration for epidemiological investigation of vaccine safety: a potential solution for low and middle-income countries? *Vaccine* 2018; **36**(3): 345–6.

163 Chen RT, Carbery B, Mac L, *et al*. The Brighton Collaboration Viral Vector Vaccines Safety Working Group (V3SWG). *Vaccine* 2015; **33**(1): 73–5.

164 Monath TP, Seligman SJ, Robertson JS, *et al*. Live virus vaccines based on a yellow fever vaccine backbone: standardized template with key considerations for a risk/benefit assessment. *Vaccine* 2015; **33**(1): 62–72.

165 Condit RC, Williamson AL, Sheets R, *et al*. Unique safety issues associated with virus-vectored vaccines: potential for and theoretical consequences of recombination with wild type virus strains. *Vaccine* 2016; **34**(51): 6610–6.

166 Dal Pan GJ, Lindquist M, Gelperin K. *Postmarketing Spontaneous Pharmacovigilance Reporting Systems: Pharmacoepidemiology*. Chichester: Wiley-Blackwell, 2012, pp. 135–57.

167 Zanoni G, Berra P, Lucchi I, *et al*. Vaccine adverse event monitoring systems across the European Union countries: time for unifying efforts. *Vaccine* 2009; **27**(25–26): 3376–84.

168 Takahashi H. Need for improved vaccine safety surveillance. *Vaccine* 2000; **19**(9–10): 1004.

169 Letourneau M, Wells G, Walop W, Duclos P. Improving global monitoring of vaccine safety: a quantitative analysis of adverse event reports in the WHO Adverse Reactions Database. *Vaccine* 2008; **26**(9): 1185–94.

170 Dey AWH, Quinn HE, Hill R, Macartney K. Surveillance of adverse events following immunisation in Australia annual report, 2014. *Commun Dis Intell* 2016; **40**(3): E377–E390.

171 Cunha MP, Dorea JG, Marques RC, Leao RS. Vaccine adverse events reported during the first ten years (1998–2008) after introduction in the state of Rondonia, Brazil. *Biomed Res Int* 2013; **2013**: 853083.

172 Top KA, Constantinescu CM, Lafleche J, *et al*. Applicability of the Brighton Collaboration Case Definition for seizure after immunization in active and passive surveillance in Canada. *Vaccine* 2013; **31**(48): 5700–5.

173 Liu D, Wu W, Li K, *et al*. Surveillance of adverse events following immunization in China: past, present, and future. *Vaccine* 2015; **33**(32): 4041–6.

174 Galindo BM, Concepcion D, Galindo MA, Perez A, Saiz J. Vaccine-related adverse events in Cuban children, 1999–2008. *MEDICC Rev* 2012; **14**(1): 38–43.

175 Aagaard L, Hansen EW, Hansen EH. Adverse events following immunization in children: retrospective analysis of spontaneous reports over a decade. *Eur J Clin Pharmacol* 2011; **67**(3): 283–8.

176 Joshi J, Das MK, Polpakara D, Aneja S, Agarwal M, Arora NK. Vaccine safety and surveillance for adverse events following immunization (AEFI) in India. *Indian J Pediatr* 2018; **85**(2): 139–48.

177 Alicino C, Merlano C, Zappettini S, *et al*. Routine surveillance of adverse events following immunization as an important tool to monitor vaccine safety. *Hum Vaccin Immunother* 2015; **11**(1): 91–4.

178 Muehlhans S, von Kleist M, Gretchukha T, *et al*. Awareness and utilization of reporting pathways for adverse events following immunization: online survey among pediatricians in Russia and Germany. *Paediatr Drugs* 2014; **16**(4): 321–30.

179 van Puijenbroek EP, Broos N, van Grootheest K. Monitoring adverse events of the vaccination campaign against influenza A (H1N1) in the Netherlands. *Drug Saf* 2010; **33**(12): 1097–108.

180 Tatley MV, Kunac DL, McNicholas A, *et al*. The Intensive Vaccines Monitoring Programme (IVMP): an electronic system to monitor vaccine safety in New Zealand. *Vaccine* 2008; **26**(22): 2746–52.

181 Schumacher Z, Bourquin C, Heininger U. Surveillance for adverse events following immunization (AEFI) in Switzerland: 1991–2001. *Vaccine* 2010; **28**(24): 4059–64.

182 Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS). *Vaccine* 2015; **33**(36): 4398–405.

183 Willis ED, Woodward M, Brown E, *et al*. Herpes zoster vaccine live: a 10-year review of post-marketing safety experience. *Vaccine* 2017; **35**(52): 7231–9.

184 Mehta U, Milstien JB, Duclos P, Folb PI. Developing a national system for dealing with adverse events following immunization. *Bull World Health Organ* 2000; **78**(2): 170–7.

185 US Department of Health and Human Services. VAERS Table of reportable events

following vaccination 2017. https://vaers.hhs.gov/docs/VAERS_Table_of_Reportable_Events_Following_Vaccination.pdf (accessed May 2019).

186 Haber P, Rashidee A, Zhou W, English R. Web-based reporting: 10 months experience in the VAERS. *Pharmacoepidemiol Drug Saf* 2003; **12**: S1–S189.

187 Centers for Disease Control and Prevention. Smallpox vaccine adverse events monitoring and response system for the first stage of the smallpox vaccination program. *Morb Mortal Wkly Rep* 2003; **52**(5): 88–9, 99.

188 Centers for Disease Control and Prevention. Influenza A (H1N1) 2009 monovalent vaccine safety monitoring: CDC planning recommendations for state, local, tribal, and territorial health officials. 2009. http://www.cdc.gov/h1n1flu/vaccination/safety_planning.htm (accessed May 2019).

189 Announcement: implementation of the Vaccine Adverse Event Reporting System 2.0 reporting form. *Morb Mortal Wkly Rep* 2017; **66**(27): 738.

190 Ribeiro-Vaz I, Silva AM, Costa Santos C, Cruz-Correia R. How to promote adverse drug reaction reports using information systems: a systematic review and meta-analysis. *BMC Med Inform Decis Mak* 2016; **16**: 27.

191 Salmon DA, Akhtar A, Mergler MJ, *et al*. Immunization-safety monitoring systems for the 2009 H1N1 monovalent influenza vaccination program. *Pediatrics* 2011; **127**(Suppl 1): S78–S86.

192 Zanardi LR, Haber P, Mootrey GT, Niu MT, Wharton M. Intussusception among recipients of rotavirus vaccine: reports to the vaccine adverse event reporting system. *Pediatrics* 2001; **107**(6): E97.

193 Centers for Disease Control and Prevention. Adverse events associated with 17D-derived yellow fever vaccination United States 2001–2002. *Morb Mortal Wkly Rep* 2002; **51**(44): 989–93.

194 Moro PL, Cragan J, Tepper N, *et al*. Enhanced surveillance of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis (Tdap) vaccines in pregnancy in the Vaccine Adverse Event Reporting System (VAERS), 2011–2015. *Vaccine* 2016; **34**(20): 2349–53.

195 Moro P, Baumblatt J, Lewis P, Cragan J, Tepper N, Cano M. Surveillance of adverse events after seasonal influenza vaccination in pregnant women and their infants in the Vaccine Adverse Event Reporting System, July 2010–May 2016. *Drug Saf* 2017; **40**(2): 145–52.

196 Centers for Disease Control and Prevention. Surveillance for adverse events associated with anthrax vaccination: U.S. Department of Defense, 1998–2000. *Morb Mortal Wkly Rep* 2000; **49**(16): 341–5.

197 Duclos P, Law B, Laflèche J, *et al*. In this supplement: vaccine safety. *CCDR* 2014; **40**: 3.

198 Labadie J, van Grootheest AC. [Adverse events following vaccination reported to the Netherlands Pharmacovigilance Center Lareb in 2004–2006]. *Ned Tijdschr Geneeskd* 2007; **151**(49): 2738–42.

199 Salisbury D, Ramsay M, Noakes K. *Immunisation against Infectious Disease*, 3rd edn. London: Stationery Office under licence from Department of Health, 2006.

200 Kim MK, Lee YK, Kim TE, Kong I, Yang HJ, Suh ES. Surveillance and compensation claims for adverse events following immunization from 2011 to 2016 in the Republic of Korea. *Clin Exp Vaccine Res* 2017; **6**(2): 146–55.

201 Chandler RE, Juhlin K, Fransson J, Caster O, Edwards IR, Noren GN. Current safety concerns with human papillomavirus vaccine: a cluster analysis of reports in VigiBase(R). *Drug Saf* 2017; **40**(1): 81–90.

202 Bettinger JA, Vanderkooi OG, MacDonald J, Kellner JD. Rapid online identification of adverse events after influenza immunization in children by PCIRN's National

Ambulatory Network. *Pediatr Infect Dis J* 2014; **33**(10): 1060–4.

203 Macartney K, Gidding HF, Trinh L, et al. Evaluation of combination measles–mumps–rubella-varicella vaccine introduction in Australia. *JAMA Pediatr* 2017; **171**(10): 992–8.

204 Thoon KC, Soh SB, Liew WK, *et al*. Active surveillance of adverse events following childhood immunization in Singapore. *Vaccine* 2014; **32**(39): 5000–5.

205 Wood RA, Berger M, Dreskin SC, *et al*. An algorithm for treatment of patients with hypersensitivity reactions after vaccines. *Pediatrics* 2008; **122**(3): e771–e777.

206 World Health Organization. *Global Vaccine Safety Blueprint: The Landscape Analysis*. Geneva: World Health Organization, 2012.

207 Graham JE, Borda-Rodriguez A, Huzair F, Zinck E. Capacity for a global vaccine safety system: the perspective of national regulatory authorities. *Vaccine* 2012; **30**(33): 4953–9.

208 World Health Organization. *Global Manual on Surveillance of Adverse Events Following Immunization*. Geneva: World Health Organization, 2014.

209 Amarasinghe A, Black S, Bonhoeffer J, *et al*. Effective vaccine safety systems in all countries: a challenge for more equitable access to immunization. *Vaccine* 2013; **31**(Suppl 2): B108–B114.

210 Lloyd JS, Milstien JB. Auto-disable syringes for immunization: issues in technology transfer. *Bull World Health Organ* 1999; **77**(12): 1001–7.

211 Lackritz ESA, Stepanchak M. *Maternal Immunization Safety Monitoring in Low- and Middle-Income Countries: A Roadmap for Program Development*. Seattle, WA: Global Alliance to Prevent Prematurity and Stillbirth (GAPPS), 2017.

212 Amanna IJ, Slifka MK. Questions regarding the safety and duration of immunity following live yellow fever vaccination. *Expert Rev Vaccines* 2016; **15**(12): 1519–33.

213 Hesseling AC, Marais BJ, Gie RP, *et al*. The risk of disseminated Bacille Calmette-Guerin (BCG) disease in HIV-infected children. *Vaccine* 2007; **25**(1): 14–18.

214 MacDonald NE, Guichard S, Amarasinghe A, Balakrishnan MR, Inter-country SEAR workshop participants. Strengthening of causality assessment of adverse events following immunization in the WHO South East Asia and Western Pacific regions: lessons from the 2014 SEAR inter-country workshop. *Vaccine* 2015; **33**(48): 6902–7.

215 Council for International Organizations of Medical Sciences. Definition and application of terms for vaccine pharmacovigilance: report of CIOMS/WHO Working Group on Vaccine Pharmacovigilance. Geneva: CIOMS, 2012.

216 Goodman M, Lamm SH, Bellman MH. Temporal relationship modeling: DTP or DT immunizations and infantile spasms. *Vaccine* 1998; **16**: 225–31.

217 Kochhar S, Bauwens J, Bonhoeffer J, for the GAIA Project Participants. Safety assessment of immunization in pregnancy. *Vaccine* 2017; **35**(48 Pt A): 6469–71.

218 Braun MM, Terracciano G, Salive ME, *et al*. Report of a US public health service workshop on hypotonic-hyporesponsive episode (HHE) after pertussis immunization. *Pediatrics* 1998; **102**(5): E52.

219 Ball R, Halsey N, Braun MM. Development of case definitions for acute encephalopathy, encephalitis, and multiple sclerosis reports to the Vaccine Adverse Reporting System. *J Clin Epidemiol* 2002; **55**: 821–30.

220 US Food and Drug Administration. International Conference on Harmonisation: Guideline on clinical safety data management: periodic safety update reports for marketed drugs. *Fed Reg* 1997; **62**: 27470–6.

221 Lateef TM, Johann-Liang R, Kaulas H, *et al*. Seizures, encephalopathy, and vaccines: experience in the National Vaccine Injury

Compensation Program. *J Pediatr* 2015; **166**(3): 576–81.

222 Goodwin H, Nash M, Gold M, Heath TC, Burgess MA. Vaccination of children following a previous hypotonic-hyporesponsive episode. *J Paediatr Child Health* 1999; **35**(6): 549–52.

223 Williams SE, Klein NP, Halsey N, *et al*. Overview of the Clinical Consult Case Review of adverse events following immunization: Clinical Immunization Safety Assessment (CISA) network 2004–2009. *Vaccine* 2011; **29**(40): 6920–7.

224 Williams SE, Pahud BA, Vellozzi C, *et al*. Causality assessment of serious neurologic adverse events following 2009 H1N1 vaccination. *Vaccine* 2011; **29**(46): 8302–8.

225 Zanoni G, Puccetti A, Dolcino M, *et al*. Dextran-specific IgG response in hypersensitivity reactions to measles–mumps–rubella vaccine. *J Allergy Clin Immunol* 2008; **122**(6): 1233–5.

226 Baxter R, Lewis N, Bohrer P, Harrington T, Aukes L, Klein NP. Sudden-onset sensorineural hearing loss after immunization: a case-centered analysis. *Otolaryngol Head Neck Surg* 2016; **155**(1): 81–6.

227 Quinn P, Gold M, Royle J, *et al*. Recurrence of extensive injection site reactions following DTPa or dTpa vaccine in children 4–6 years old. *Vaccine* 2011; **29**(25): 4230–7.

228 Klein NP, Edwards KM, Sparks RC, Dekker CL. Recurrent sterile abscesses following aluminium adjuvant-containing vaccines. *BMJ Case Rep* 2009; pii: bcr09.2008.0951.

229 Halsey NA, Edwards KM, Dekker CL, *et al*. Algorithm to assess causality after individual adverse events following immunizations. *Vaccine* 2012; **30**(39): 5791–8.

230 Kramer MS, Lane DA. Causal propositions in clinical research and practice. *J Clin Epidemiol* 1992; **45**(6): 639–49.

231 Stratton KR, Howe CJ, Johnston RB, Jr. Adverse events associated with childhood vaccines other than pertussis and rubella. Summary of a report from the Institute of Medicine. *JAMA* 1994; **271**(20): 1602–5.

232 Hill AB. The environment and disease: association or causation. *Proc Roy Soc Med* 1965; **58**: 295–300.

233 Mascolo A, Scavone C, Sessa M, *et al*. Can causality assessment fulfill the new European definition of adverse drug reaction? A review of methods used in spontaneous reporting. *Pharmacol Res* 2017; **123**: 122–9.

234 Jones JK. Assessing causality of case reports of suspected adverse events. In Strom BL, Kimmel SE, Hennessy S, eds., *Pharmacoepidemiology*, 5th edn. Chichester: John Wiley & Sons, 2011, pp. 583–600.

235 Edsall G, Elliott MW, Peebles TC, Eldred MC. Excessive use of tetanus toxoid boosters. *JAMA* 1967; **202**(1): 111–13.

236 Institute of Medicine (US) Committee to Review the Adverse Consequences of Pertussis and Rubella Vaccines; Howson CP, Howe CJ, Fineberg HV, eds. Adverse effects of pertussis and rubella vaccines: a report of the Committee to Review the Adverse Consequences of Pertussis and Rubella Vaccines. Washington, DC: National Academies Press, 1991.

237 Centers for Disease Control and Prevention. Rubella vaccination during pregnancy: United States, 1971–1988. *Morb Mortal Wkly Rep* 1989; **38**(17): 289–93.

238 Duffy J, Weintraub E, Hambidge SJ, *et al*. Febrile seizure risk after vaccination in children 6 to 23 months. *Pediatrics* 2016; **138**(1): pii: e20160320.

239 Collet JP, MacDonald N, Cashman N, Pless R. Monitoring signals for vaccine safety: the assessment of individual adverse event reports by an expert advisory committee. Advisory Committee on Causality Assessment. *Bull World Health Organ* 2000; **78**(2): 178–85.

240 Williams SE, Pahud BA, Vellozzi C, *et al*. Causality assessment of serious neurologic

adverse events following 2009 H1N1 vaccination. *Vaccine* 2011; **29**(46): 8302–8.

241 World Health Organization. *Causality Assessment of an Adverse Event Following Immunization (AEFI): User Manual for the Revised WHO Classification*, 2nd edn. Geneva: World Health Organization, 2018.

242 Puliyel J, Naik P. Revised World Health Organization (WHO)'s causality assessment of adverse events following immunization: a critique. *F1000Res* 2018; **7**: 243.

243 Rennels MB, Parashar UD, Holman RC, Le CT, Chang HG, Glass RI. Lack of an apparent association between intussusception and wild or vaccine rotavirus infection. *Pediatr Infect Dis J* 1998; **17**(10): 924–5.

244 Domingo C, Niedrig M. Safety of 17D derived yellow fever vaccines. *Expert Opin Drug Saf* 2009; **8**(2): 211–21.

245 Monath TP. Suspected yellow fever vaccine-associated viscerotropic adverse events (1973 and 1978), United States. *Am J Trop Med Hyg* 2010; **82**(5): 919–21.

246 Skowronski DM, Strauss B, De Serres G, *et al*. Oculo-respiratory syndrome: a new influenza vaccine-associated adverse event? *Clin Infect Dis* 2003; **36**(6): 705–13.

247 Neale T. ACIP: warning for Guillain-Barré Syndrome lifted. *MedPage Today*, **24** June 2010.

248 Leroy Z, Broder K, Menschik D, Shimabukuro T, Martin D. Febrile seizures after 2010–2011 influenza vaccine in young children, United States: a vaccine safety signal from the Vaccine Adverse Event Reporting System. *Vaccine* 2012; **30**(11): 2020–3.

249 Tse A, Tseng HF, Greene SK, Vellozzi C, Lee GM, Group VSDRCAIW. Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink Project, 2010–2011. *Vaccine* 2012; **30**(11): 2024–31.

250 Carson JL, Strom BL, Maislin G. Screening for unknown effects of newly marketed drugs. In Stom BL, ed., *Pharmacoepidemiology*. Chichester: John Wiley & Sons, 1994, p. 431–47.

251 Niu MT, Erwin DE, Braun MM. Data mining in the US Vaccine Adverse Event Reporting System (VAERS): early detection of intussusception and other events after rotavirus vaccination. *Vaccine* 2001; **19**(32): 4627–34.

252 DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 1999; **53**(3): 177–90.

253 Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf* 2002; **25**(6): 381–92.

254 Pool V, Chen RT. Association rule discovery as a signal generation tool for the Vaccine Adverse Event Reporting System. *Pharmacoepidemiol Drug Saf* 2003; **11**: S57.

255 Rolka H, Bracy D, Russell C, Fram D, Ball R. Using simulation to assess the sensitivity and specificity of a signal detection tool for multidimensional public health surveillance data. *Stat Med* 2005; **24**(4): 551–62.

256 Rosenthal S, Chen R, Hadler S. The safety of acellular pertussis vaccine vs whole-cell pertussis vaccine: a postmarketing assessment. *Arch Pediatr Adolesc Med* 1996; **150**(5): 457–60.

257 Ball R. Methods for ensuring vaccine safety. *Exp Rev Vaccines* 2002; **1**: 161–8.

258 Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001; **10**(6): 483–6.

259 Banks D, Woo EJ, Burwen DR, Perucci P, Braun MM, Ball R. Comparing data mining methods on the VAERS database. *Pharmacoepidemiol Drug Saf* 2005; **14**(9): 601–9.

260 Davis RL. Vaccine safety surveillance systems: critical elements and lessons learned in the development of the US Vaccine Safety Datalink's rapid cycle analysis capabilities. *Pharmaceutics* 2013; **5**(1): 168–78.

261 Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T. The Vaccine Safety Datalink: a model for monitoring immunization safety. *Pediatrics* 2011; **127**(Suppl 1): S45–S53.

262 Klein NP, Fireman B, Yih WK, *et al*. Measles–mumps–rubella–varicella combination vaccine and the risk of febrile seizures. *Pediatrics* 2010; **126**(1): e1–e8.

263 Yih WK, Maro JC, Dashevsky I, *et al*. Sentinel CBER Surveillance Protocol: evaluation of HPV9 (Gardasil 9) vaccine safety surveillance using the TREESCAN data mining method. Washington, DC: Sentinel Initiative, 2016.

264 Balinska MA. L'affaire hepatite B en France. *Espirit* 2001; **276**: 34–48.

265 da Cunha SS, Rodrigues LC, Barreto ML, Dourado I. Outbreak of aseptic meningitis and mumps after mass vaccination with MMR vaccine using the Leningrad-Zagreb mumps strain. *Vaccine* 2002; **20**(7–8): 1106–12.

266 Uhari M, Rantala H, Niemela M. Cluster of childhood Guillain-Barré cases after an oral poliovaccine campaign. *Lancet* 1989; **2**(8660): 440–1.

267 Berg SW, Mitchell BS, Hanson RK, *et al*. Systemic reactions in U.S. Marine Corps personnel who received Japanese encephalitis vaccine. *Clin Infect Dis* 1997; **24**(2): 265–6.

268 Kilroy AW, Schaffner W, Fleet WF, Jr., Lefkowitz LB, Jr., Karzon DT, Fenichel GM. Two syndromes following rubella immunization: clinical observations and epidemiological studies. *JAMA* 1970; **214**(13): 2287–92.

269 Vannice KS, Keita M, Sow SO, *et al*. Active surveillance for adverse events after a mass vaccination campaign with a group A meningococcal conjugate vaccine (PsA-TT) in Mali. *Clin Infect Dis* 2015; **61**(Suppl 5): S493–S500.

270 da Silveira CM, Salisbury DM, de Quadros CA. Measles vaccination and Guillain-Barré syndrome. *Lancet* 1997; **349**(9044): 14–16.

271 Rantala H, Cherry JD, Shields WD, Uhari M. Epidemiology of Guillain-Barré syndrome in children: relationship of oral polio vaccine administration to occurrence. *J Pediatr* 1994; **124**(2): 220–3.

272 Sexton K, McNicholas A, Galloway Y, *et al*. Henoch-Schonlein purpura and meningococcal B vaccination. *Arch Dis Child* 2009; **94**(3): 224–6.

273 Yang TU, Kim HJ, Lee YK, Park YJ. Psychogenic illness following vaccination: exploratory study of mass vaccination against pandemic influenza A (H1N1) in 2009 in South Korea. *Clin Exp Vaccine Res* 2017; **6**(1): 31–7.

274 Huang WT, Chuang JH, Kuo SH. Monitoring the safety of pandemic H1N1 vaccine. *Lancet* 2010; **375**(9721): 1164.

275 Centers for Disease Control and Prevention. Preliminary results: surveillance for Guillain-Barré syndrome after receipt of influenza A (H1N1) 2009 monovalent vaccine: United States, 2009–2010. *Morb Mortal Wkly Rep* 2010; **59**(21): 657–61.

276 Centers for Disease Contro and Prevention. Safety of influenza A (H1N1) 2009 monovalent vaccines: United States, October 1–November 24, 2009. *Morb Mortal Wkly Rep* 2009; **58**(48): 1351–6.

277 Lee GM, Greene SK, Weintraub ES, *et al*. H1N1 and seasonal influenza vaccine safety in the Vaccine Safety Datalink project. *Am J Prev Med* 2011; **41**(2): 121–8.

278 Black S, Eskola J, Siegrist CA, *et al*. Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. *Lancet* 2009; **374**(9707): 2115–22.

279 Schatz M, Chambers CD, Jones KL, Louik C, Mitchell AA. Safety of influenza immunizations and treatment during pregnancy: the Vaccines and Medications in Pregnancy Surveillance System. *Am J Obstet Gynecol* 2011; **204**(6 Suppl 1): S64–S68.

280 Weibel D, Sturkenboom M, Black S, *et al.* Narcolepsy and adjuvanted pandemic influenza A (H1N1) 2009 vaccines - Multi-country assessment. *Vaccine* 2018; **36**(41): 6202–11.

281 Braun MM, Ellenberg SS. Descriptive epidemiology of adverse events after immunization: reports to the Vaccine Adverse Event Reporting System (VAERS), 1991–1994. *J Pediatr* 1997; **131**(4): 529–35.

282 Singleton JA, Lloyd JC, Mootrey GT, Salive ME, Chen RT. An overview of the Vaccine Adverse Event Reporting System (VAERS) as a surveillance system. VAERS Working Group. *Vaccine* 1999; **17**(22): 2908–17.

283 Zhou W, Pool V, Iskander JK, *et al.* Surveillance for safety after immunization: Vaccine Adverse Event Reporting System (VAERS) – United States, 1991–2001. *MMWR Surveill Summ* 2003; **52**(1): 1–24.

284 Varricchio F, Iskander J, DeStefano F, Pless R, Braun MM, Chen RT. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. *Pediatr Infect Dis J* 2004; **23**: 287–94.

285 Beeler J, Varricchio F, Wise R. Thrombocytopenia after immunization with measles vaccines: review of the Vaccine Adverse Events Reporting System (1990 to 1994). *Pediatr Infect Dis J* 1996; **15**(1): 88–90.

286 Pool V, Braun MM, Kelso JM, *et al.* Prevalence of anti-gelatin IgE antibodies in people with anaphylaxis after measles-mumps rubella vaccine in the United States. *Pediatrics* 2002; **110**(6): e71.

287 Braun MM, Patriarca PA, Ellenberg SS. Syncope after immunization. *Arch Pediatr Adolesc Med* 1997; **151**(3): 255–9.

288 Haber P, Chen RT, Zanardi LR, Mootrey GT, English R, Braun MM. An analysis of rotavirus vaccine reports to the vaccine adverse event reporting system: more than intussusception alone? *Pediatrics* 2004; **113**(4): e353–e9.

289 Wise RP, Salive ME, Braun MM, *et al.* Postlicensure safety surveillance for varicella vaccine. *JAMA* 2000; **284**(10): 1271–9.

290 Slade BA, Leidel L, Vellozzi C, *et al.* Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. *JAMA* 2009; **302**(7): 750–7.

291 Woo EJ, Moro PL, Cano M, Jankosky C. Postmarketing safety surveillance of trivalent recombinant influenza vaccine: reports to the Vaccine Adverse Event Reporting System. *Vaccine* 2017; **35**(42): 5618–21.

292 Moro PL, Winiecki S, Lewis P, Shimabukuro TT, Cano M. Surveillance of adverse events after the first trivalent inactivated influenza vaccine produced in mammalian cell culture (Flucelvax(R)) reported to the Vaccine Adverse Event Reporting System (VAERS), United States, 2013–2015. *Vaccine* 2015; **33**(48): 6684–8.

293 Centers for Disease Control and Prevention. Updated recommendations for use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine (Tdap) in pregnant women: Advisory Committee on Immunization Practices (ACIP), 2012. *Morb Mortal Wkly Rep* 2013; **62**(7): 131–5.

294 Hibbs BF, Moro PL, Lewis P, Miller ER, Shimabukuro TT. Vaccination errors reported to the Vaccine Adverse Event Reporting System, (VAERS) United States, 2000–2013. *Vaccine* 2015; **33**(28): 3171–8.

295 Su JR, Miller ER, Duffy J, Baer BM, Cano MV. Notes from the field: administration error involving a meningococcal conjugate vaccine – United States, March 1, 2010–September 22, 2015. *Morb Mortal Wkly Rep* 2016; **65**(6): 161–2.

296 Vellozzi C, Broder KR, Haber P, *et al.* Adverse events following influenza A (H1N1)

2009 monovalent vaccines reported to the Vaccine Adverse Event Reporting System, United States, October 1, 2009–January 31, 2010. *Vaccine* 2010; **28**(45): 7248–55.

297 Dayan GH, Iskander J, Glasser J, English-Bullard R, Fullerton KE, Chen R. Tracking vaccine lot lifecycles using reports to the Vaccine Adverse Event Reporting System (VAERS). *Pharmacoepidemiol Drug Saf* 2005; **14**(10): 671–6.

298 Vellozzi C, Mitchell T, Miller E, Casey CG, Eidex RB, Hayes EB. Yellow fever vaccine-associated viscerotropic disease (YEL-AVD) and corticosteroid therapy: eleven United States cases, 1996–2004. *Am J Trop Med Hyg* 2006; **75**(2): 333–6.

299 Barwick R. History of thymoma and yellow fever vaccination. *Lancet* 2004; **364**(9438): 936.

300 Chaves SS, Haber P, Walton K, *et al*. Safety of varicella vaccine after licensure in the United States: experience from reports to the Vaccine Adverse Event Reporting System, 1995-2005. *J Infect Dis* 2008; **197**(Suppl 2): S170–S177.

301 Huang WT, Suh C, Campagna E, *et al*. Adherence to the Advisory Committee on Immunization Practices recommendation to prevent injuries from postvaccination syncope: a national physician survey. *Am J Prev Med* 2011; **41**(3): 317–21.

302 Rosenthal S, Chen R. The reporting sensitivities of two passive surveillance systems for vaccine adverse events. *Am J Public Health* 1995; **85**(12): 1706–9.

303 Verstraeten T, Baughman AL, Cadwell B, Zanardi L, Haber P, Chen RT. Enhancing vaccine safety surveillance: a capture–recapture analysis of intussusception after rotavirus vaccination. *Am J Epidemiol* 2001; **154**(11): 1006–12.

304 Chen RT, Kent JH, Rhodes PH, *et al*. Investigation of a possible association between influenza vaccination and Guillain-Barré syndrome in the United States, 1990–1991. *Postmarket Surveill* 1992; **6**: 5–6.

305 Chen RT. Evaluation of vaccine safety after the events of 11 September 2001: role of cohort and case–control studies. *Vaccine* 2004; **22**(15–16): 2047–53.

306 De Serres G, Toth E, Menard S, *et al*. Oculo-respiratory syndrome after influenza vaccination: trends over four influenza seasons. *Vaccine* 2005; **23**(28): 3726–32.

307 Spila-Alegiani S, Salmaso S, Rota MC, Tozzi AE, Raschetti R. Reactogenicity in the elderly of nine commercial influenza vaccines: results from the Italian SVEVA study. Study for the evaluation of adverse events of influenza vaccination. *Vaccine* 1999; **17**(15–16): 1898–904.

308 Niu MT, Rhodes P, Salive M, *et al*. Comparative safety of two recombinant hepatitis B vaccines in children: data from the Vaccine Adverse Event Reporting System (VAERS) and Vaccine Safety Datalink (VSD). *J Clin Epidemiol* 1998; **51**(6): 503–10.

309 Food and Drug Administration. Postmarketing expedited adverse experience reporting for human drug and licensed biological products; increased frequency reports. *Fed Reg* 1997; **62**: 34166–8.

310 Murthy N, Rodgers L, Pabst L, Fiebelkorn AP, Ng T. Progress in childhood vaccination data in immunization information systems: United States, 2013–2016. *Morb Mortal Wkly Rep* 2017; **66**(43): 1178–81.

311 Gessner BD, Knobel DL, Conan A, Finn A. Could the RTS,S/AS01 meningitis safety signal really be a protective effect of rabies vaccine? *Vaccine* 2017; **35**(5): 716–21.

312 Sano J, Chaitman BR, Swindle J, Frey SE. Electrocardiography screening for cardiotoxicity after modified Vaccinia Ankara vaccination. *Am J Med* 2009; **122**(1): 79–84.

313 Enwere GC, Paranjape G, Kulkarni PS, *et al*. Safety monitoring in group A meningococcal conjugate vaccine trials: description, challenges, and lessons. *Clin Infect Dis* 2015; **61**(Suppl 5): S501–S506.

314 Patriarca PA, Laender F, Palmeira G, *et al.* Randomised trial of alternative formulations of oral poliovaccine in Brazil. *Lancet* 1988; **1**(8583): 429–33.

315 Bhargava I, Chhaparwal BC, Phadke MA, Irani SF, Chakladhar BK, Maheshwari CP. Reactogenecity of indigenously produced measles vaccine. *Indian Pediatr* 1996; **33**(10): 827–31.

316 Orenstein WA, Markowitz L, Preblud SR, Hinman AR, Tomasi A, Bart KJ. Appropriate age for measles vaccination in the United States. *Dev Biol Stand* 1986; **65**: 13–21.

317 Booy R, Taylor SA, Dobson SR, *et al.* Immunogenicity and safety of PRP-T conjugate vaccine given according to the British accelerated immunisation schedule. *Arch Dis Child* 1992; **67**(4): 475–8.

318 Deforest A, Long SS, Lischner HW, *et al.* Simultaneous administration of measles–mumps–rubella vaccine with booster doses of diphtheria–tetanus–pertussis and poliovirus vaccines. *Pediatrics* 1988; **81**(2): 237–46.

319 Scheifele D, Law B, Mitchell L, Ochnio J. Study of booster doses of two Haemophilus influenzae type b conjugate vaccines including their interchangeability. *Vaccine* 1996; **14**(15): 1399–406.

320 Aaby P, Samb B, Simondon F, *et al.* A comparison of vaccine efficacy and mortality during routine use of high-titre Edmonston-Zagreb and Schwarz standard measles vaccines in rural Senegal. *Trans R Soc Trop Med Hyg* 1996; **90**(3): 326–30.

321 Extended Programme on Immunization. Safety of high titer measles vaccines. *Wkly Epidemiol Rec* 1992; **67**: 357–61.

322 Black SB, Shinefield HR. b-CAPSA I Haemophilus influenzae, type b, capsular polysaccharide vaccine safety. *Pediatrics* 1987; **79**(3): 321–5.

323 Vadheim CM, Greenberg DP, Marcy SM, Froeschle J, Ward JI. Safety evaluation of PRP-D Haemophilus influenzae type b conjugate vaccine in children immunized at 18 months of age and older: follow-up study of 30,000 children. *Pediatr Infect Dis J* 1990; **9**(8): 555–61.

324 Chen RT. Safety of acellular pertussis vaccine: follow-up studies. *Dev Biol Stand* 1997; **89**: 373–5.

325 Centers for Disease Control and Prevention. Unintentional administration of varicella virus vaccine: United States, 1996. *Morb Mortal Wkly Rep* 1996; **45**(46): 1017–18.

326 Black S, Shinefield H, Ray P, *et al.* Postmarketing evaluation of the safety and effectiveness of varicella vaccine. *Pediatr Infect Dis J* 1999; **18**(12): 1041–6.

327 Wang Y, Dong D, Cheng G, Zuo S, Liu D, Du X. Post-marketing surveillance of live-attenuated Japanese encephalitis vaccine safety in China. *Vaccine* 2014; **32**(44): 5875–9.

328 Liu ZL, Hennessy S, Strom BL, *et al.* Short-term safety of live attenuated Japanese encephalitis vaccine (SA14-14-2): results of a randomized trial with 26,239 subjects. *J Infect Dis* 1997; **176**(5): 1366–9.

329 Henderson DA, Witte JJ, Morris L, *et al.* Paralytic disease associated with oral polio vaccines. *JAMA* 1964; **190**: 153–60.

330 Simon PA, Chen RT, Elliott JA, Schwartz B. Outbreak of pyogenic abscesses after diphtheria and tetanus toxoids and pertussis vaccination. *Pediatr Infect Dis J* 1993; **12**(5): 368–71.

331 Walker AM, Jick H, Perera DR, Thompson RS, Knauss TA. Diphtheria–tetanus–pertussis immunization and sudden infant death syndrome. *Am J Public Health* 1987; **77**(8): 945–51.

332 Griffin MR, Ray WA, Livengood JR, Schaffner W. Risk of sudden infant death syndrome after immunization with the diphtheria-tetanus-pertussis vaccine. *N Engl J Med* 1988; **319**(10): 618–23.

333 Griffin MR, Ray WA, Mortimer EA, Fenichel GM, Schaffner W. Risk of seizures and

encephalopathy after immunization with the diphtheria–tetanus–pertussis vaccine. *JAMA* 1990; **263**(12): 1641–5.

334 Chen RT, Glasser JW, Rhodes PH, *et al*. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. The Vaccine Safety Datalink Team. *Pediatrics* 1997; **99**(6): 765–73.

335 Mayes R. The origins, development, and passage of Medicare's revolutionary prospective payment system. *J Hist Med Allied Sci* 2007; **62**(1): 21–55.

336 Fahey KR. The pioneering role of the Vaccine Safety Datalink Project (VSD) to advance collaborative research and distributed data networks. *EGEMS (Wash DC)* 2015; **3**(1): 1195.

337 DeStefano F. The Vaccine Safety Datalink project. *Pharmacoepidemiol Drug Saf* 2001; **10**(5): 403–6.

338 Ray P, Black S, Shinefield H, *et al*. Risk of chronic arthropathy among women after rubella vaccination. Vaccine Safety Datalink Team. *JAMA* 1997; **278**(7): 551–6.

339 Thompson WW, Price C, Goodson B, *et al*. Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years. *N Engl J Med* 2007; **357**(13): 1281–92.

340 France EK, Glanz J, Xu S, *et al.* Risk of immune thrombocytopenic purpura after measles–mumps–rubella immunization in children. *Pediatrics* 2008; **121**(3): e687–e692.

341 Davis RL, Marcuse E, Black S, *et al*. MMR2 immunization at 4 to 5 years and 10 to 12 years of age: a comparison of adverse clinical events after immunization in the Vaccine Safety Datalink project. The Vaccine Safety Datalink Team. *Pediatrics* 1997; **100**(5): 767–71.

342 Sukumaran L, McCarthy NL, Kharbanda EO, *et al*. Infant hospitalizations and mortality after maternal vaccination. *Pediatrics* 2018; **141**(3): pii: e20173310.

343 Gee J, Naleway A, Shui I, *et al*. Monitoring the safety of quadrivalent human papillomavirus vaccine: findings from the Vaccine Safety Datalink. *Vaccine* 2011; **29**(46): 8279–84.

344 Hambidge SJ, Glanz JM, France EK, *et al*. Safety of trivalent inactivated influenza vaccine in children 6 to 23 months old. *JAMA* 2006; **296**(16): 1990–7.

345 Lin ND, Kleinman K, Chan KA, *et al*. Multiple vaccinations and the risk of medically attended fever. *Vaccine* 2010; **28**(25): 4169–74.

346 Daley MF, Yih WK, Glanz JM, *et al*. Safety of diphtheria, tetanus, acellular pertussis and inactivated poliovirus (DTaP-IPV) vaccine. *Vaccine* 2014; **32**(25): 3019–24.

347 Nelson JC, Yu O, Dominguez-Islas CP, *et al*. Adapting group sequential methods to observational postlicensure vaccine safety surveillance: results of a pentavalent combination DTaP-IPV-Hib vaccine safety study. *Am J Epidemiol* 2013; **177**(2): 131–41.

348 Cohn A, Rodewald LE, Orenstein WA, Schuchat A. Immunizations in the United States. In Plotkin SA, Orenstein WA, Offit PA, Edwards KM, eds., *Plotkin's Vaccines*, 7th edn, Philadelphia, PA: Elsevier, 2017, pp. 1421–40.

349 Verstraeten T, Jumaan AO, Mullooly JP, *et al*. A retrospective cohort study of the association of varicella vaccine failure with asthma, steroid use, age at vaccination, and measles–mumps–rubella vaccination. *Pediatrics* 2003; **112**(2): e98–e103.

350 Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Stat Med* 2010; **29**(2): 284–95.

351 Lieu TA, Kulldorff M, Davis RL, *et al*. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45**(10 Supl 2): S89–S95.

352 Narwaney KJ, Breslin K, Ross CA, *et al*. Vaccine adverse events in a safety net healthcare system and a managed care organization. *Vaccine* 2017; **35**(9): 1335–40.

353 Rothman KJ. *Causal Inference*. Chestnut Hill, MA: Epidemiology Resources, 1988.

354 Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 291–7.

355 Baker MA, Nguyen M, Cole DV, Lee GM, Lieu TA. Post-licensure Rapid Immunization Safety Monitoring program (PRISM) data characterization. *Vaccine* 2013; **31**(Suppl 10): K98–K112.

356 Yih WK, Lee GM, Lieu TA, *et al*. Surveillance for adverse events following receipt of pandemic 2009 H1N1 vaccine in the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) System, 2009–2010. *Am J Epidemiol* 2012; **175**(11): 1120–8.

357 Kawai AT, Martin D, Kulldorff M, *et al*. Febrile seizures after 2010-2011 trivalent inactivated influenza vaccine. *Pediatrics* 2015; **136**(4): e848–e855.

358 Yih WK, Greene SK, Zichittella L, *et al*. Evaluation of the risk of venous thromboembolism after quadrivalent human papillomavirus vaccination among US females. *Vaccine* 2016; **34**(1): 172–8.

359 Yih WK, Lieu TA, Kulldorff M, *et al*. Intussusception risk after rotavirus vaccination in U.S. infants. *N Engl J Med* 2014; **370**(6): 503–12.

360 Kulldorff M, Dashevsky I, Avery TR, *et al*. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf* 2013; **22**(5): 517–23.

361 Hanf M, Quantin C, Farrington P, *et al*. Validation of the French national health insurance information system as a tool in vaccine safety assessment: application to febrile convulsions after pediatric measles/mumps/rubella immunization. *Vaccine* 2013; **31**(49): 5856–62.

362 Schink T, Holstiege J, Kowalzik F, Zepp F, Garbe E. Risk of febrile convulsions after MMRV vaccination in comparison to MMR or MMR+V vaccination. *Vaccine* 2014; **32**(6): 645–50.

363 MacDonald SE, Dover DC, Simmonds KA, Svenson LW. Risk of febrile seizures after first dose of measles–mumps–rubella–varicella vaccine: a population-based cohort study. *CMAJ* 2014; **186**(11): 824–9.

364 Kurz X, Bauchau V, Mahy P, *et al*. The ADVANCE Code of Conduct for collaborative vaccine studies. *Vaccine* 2017; **35**(15): 1844–55.

365 Ali M, Rath B, Thiem VD. Vaccine safety monitoring systems in developing countries: an example of the Vietnam model. *Curr Drug Saf* 2015; **10**(1): 60–7.

366 Black S. Global Vaccine Safety DataNet meeting. *Expert Rev Vaccines* 2008; **7**(1): 15–20.

367 Sankoh O, Byass P. *The INDEPTH Network: Filling Vital Gaps in Global Epidemiology*. Oxford: Oxford University Press, 2012.

368 Meeting of the Strategic Advisory Group of Experts on Immunization, October 2015: conclusions and recommendations. *Wkly Epidemiol Rec* 2015; **90**(50): 681–99.

369 Cashman P, Macartney K, Khandaker G, King C, Gold M, Durrheim DN. Participant-centred active surveillance of adverse events following immunisation: a narrative review. *Int Health* 2017; **9**(3): 164–76.

370 Crawford NW, Clothier H, Hodgson K, Selvaraj G, Easton ML, Buttery JP. Active surveillance for adverse events following immunization. *Expert Rev Vaccines* 2014; **13**(2): 265–76.

371 Looker C, Kelly H. No-fault compensation following adverse events attributed to vaccination: a review of international programmes. *Bull World Health Organ* 2011; **89**(5): 371–8.

372 Cook KM, Evans G. The National Vaccine Injury Compensation Program. *Pediatrics* 2011; **127**(Suppl 1): S74–S77.

373 Mullooly J, Drew L, DeStefano F, *et al*. Quality of HMO vaccination databases used to monitor childhood vaccine safety. Vaccine Safety DataLink Team. *Am J Epidemiol* 1999; **149**(2): 186–94.

374 Greene SK, Shi P, Dutta-Linn MM, *et al*. Accuracy of data on influenza vaccination status at four Vaccine Safety Datalink sites. *Am J Prev Med* 2009; **37**(6): 552–5.

375 Wilton R, Pennisi A. Evaluating the accuracy of transcribed computer-stored immunization data. *Pediatrics* 1994; **94**: 902–6.

376 Australian Government Department of Human Services. Australian Immunisation Register. https://www.humanservices.gov.au/individuals/services/medicare/australian-immunisation-register (accessed May 2019).

377 Health Canada. Immunization Registries. https://www.canada.ca/en/health-canada/services/healthy-living/your-health/medical-information/immunization-registries.html (accessed May 2019).

378 Kang G, Zhou M, Wang Z, Yu J, Tang F. Ten years of experience and progress of electronic immunization registry system in Jiangsu Province, China. *Hum Vaccin Immunother* 2018; **14**(1): 163–4.

379 Danovaro-Holliday MC, Ortiz C, Cochi S, Ruiz-Matus C. Electronic immunization registries in Latin America: progress and lessons learned. *Rev Panam Salud Publica* 2014; **35**(5–6): 453–7.

380 Amirthalingam G, White J, Ramsay M. Measuring childhood vaccine coverage in England: the role of Child Health Information Systems. *Euro Surveill* 2012; **17**(16): pii: 20149.

381 Wilson K, Atkinson KM, Deeks SL, Crowcroft NS. Improving vaccine registries through mobile technologies: a vision for mobile enhanced immunization information systems. *J Am Med Inform Assoc* 2016; **23**(1): 207–11.

382 Gibson DG, Ochieng B, Kagucia EW, *et al*. Mobile phone-delivered reminders and incentives to improve childhood immunisation coverage and timeliness in Kenya (M-SIMU): a cluster randomised controlled trial. *Lancet Glob Health* 2017; **5**(4): e428–e438.

383 McMahon BJ, Helminiak C, Wainwright RB, Bulkow L, Trimble BA, Wainwright K. Frequency of adverse reactions to hepatitis B vaccine in 43,618 persons. *Am J Med* 1992; **92**(3): 254–6.

384 Daily A, Kennedy ED, Fierro LA, *et al*. Evaluation of scanning 2D barcoded vaccines to improve data accuracy of vaccines administered. *Vaccine* 2016; **34**(47): 5802–7.

385 Global Advisory Committee on Vaccine Safety, 3–4 December 2009. *Wkly Epidemiol Rec* 2010; **85**(5): 29–33.

386 DeStefano F, Verstraeten T, Jackson LA, *et al*. Vaccinations and risk of central nervous system demyelinating diseases in adults. *Arch Neurol* 2003; **60**(4): 504–9.

387 Poser CM. Neurological complications of swine influenza vaccination. *Acta Neurol Scand* 1982; **66**(4): 413–31.

388 Hotopf M, David A, Hull L, Ismail K, Unwin C, Wessely S. Role of vaccinations as risk factors for ill health in veterans of the Gulf war: cross sectional study. *BMJ* 2000; **320**(7246): 1363–7.

389 Greene SK, Kulldorff M, Lewis EM, *et al*. Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *Am J Epidemiol* 2010; **171**(2): 177–88.

390 Rothman KJ, Greenland S, Lash T. *Modern Epidemiology*. New York: Lippincott Williams & Wilkins, 2008.

391 Szklo M, Nieto J. Basic study designs in analytical epidemiology. In Szklo M, Nieto J, eds., *Epidemiology: Beyond the Basics*, 2nd edn. Sudbury, MA: Jones and Bartlett, 2006, pp. 3–43.

392 Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**(11): 1165–73.

393 Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ* 2016; **354**: i4515.

394 Farrington CP, Miller E, Taylor B. MMR and autism: further evidence against a causal association. *Vaccine* 2001; **19**(27): 3632–5.

395 Kramarz P, DeStefano F, Gargiullo PM, *et al*. Does influenza vaccination exacerbate asthma? Analysis of a large cohort of children with asthma. Vaccine Safety Datalink Team. *Arch Fam Med* 2000; **9**(7): 617–23.

396 Farrington CP, Whitaker HJ. Semiparametric analysis of case series data. *Appl Stat* 2006; **55**(5): 553–94.

397 Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009; **10**(1): 3–16.

398 Glanz JM, McClure DL, Xu S, *et al*. Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *J Clin Epidemiol* 2006; **59**(8): 808–18.

399 Fireman B, Lee J, Lewis N, Bembom O, van der LM, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. *Am J Epidemiol* 2009; **170**(5): 650–6.

400 Qian L, Tseng HF, Sy LS, Jacobsen SJ. Confounder adjustment in vaccine safety studies: comparing three offset terms for case-centered approach. *Vaccine* 2013; **31**(2): 431–5.

401 Greene SK, Rett MD, Vellozzi C, *et al*. Guillain-Barré Syndrome, influenza vaccination, and antecedent respiratory and gastrointestinal infections: a case-centered analysis in the Vaccine Safety Datalink, 2009–2011. *PLoS One* 2013; **8**(6): e67185.

402 Ray GT, Lewis N, Goddard K, *et al*. Asthma exacerbations among asthmatic children receiving live attenuated versus inactivated influenza vaccines. *Vaccine* 2017; **35**(20): 2668–75.

403 Baxter R, Lewis E, Fireman B, DeStefano F, Gee J, Klein NP. Case-centered analysis of optic neuritis after vaccines. *Clin Infect Dis* 2016; **63**(1): 79–81.

404 Rowhani-Rahbar A, Fireman B, Lewis E, *et al*. Effect of age on the risk of fever and seizures following immunization with measles-containing vaccines in children. *JAMA Pediatr* 2013; **167**(12): 1111–17.

405 Cheng DR, Perrett KP, Choo S, Danchin M, Buttery JP, Crawford NW. Pediatric anaphylactic adverse events following immunization in Victoria, Australia from 2007 to 2013. *Vaccine* 2015; **33**(13): 1602–7.

406 Oberle D, Pavel J, Rieck T, *et al*. Anaphylaxis after immunization of children and adolescents in Germany. *Pediatr Infect Dis J* 2016; **35**(5): 535–41.

407 McNeil MM, Weintraub ES, Duffy J, *et al*. Risk of anaphylaxis after vaccination in children and adults. *J Allergy Clin Immunol* 2016; **137**(3): 868–78.

408 Arnoldussen DL, Linehan M, Sheikh A. BCG vaccination and allergy: a systematic review and meta-analysis. *J Allergy Clin Immunol* 2011; **127**(1): 246–53, e1–21.

409 Martin Arias LH, Sanz R, Sainz M, Treceno C, Carvajal A. Guillain-Barré syndrome and influenza vaccines: a meta-analysis. *Vaccine* 2015; **33**(31): 3773–8.

410 Dodd CN, Romio SA, Black S, *et al*. International collaboration to assess the risk of Guillain Barré Syndrome following influenza A (H1N1) 2009 monovalent vaccines. *Vaccine* 2013; **31**(40): 4448–58.

411 Xu S, Zhang L, Nelson J, *et al*. Identifying optimal risk windows for self-controlled case series studies of vaccine safety. *Stat Med* 2011; **30**(7): 742–52.

412 Gellin BG, Maibach EW, Marcuse EK. Do parents understand immunizations? A national telephone survey. *Pediatrics* 2000; **106**(5): 1097–102.

413 Gust DA, Darling N, Kennedy A, Schwartz B. Parents with doubts about vaccines: which

vaccines and reasons why. *Pediatrics* 2008; **122**(4): 718–25.

414 Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM. Parental vaccine safety concerns in 2009. *Pediatrics* 2010; **125**(4): 654–9.

415 Dempsey AF, Schaffer S, Singer D, Butchart A, Davis M, Freed GL. Alternative vaccination schedule preferences among parents of young children. *Pediatrics* 2011; **128**(5): 848–56.

416 Kennedy A, Basket M, Sheedy K. Vaccine attitudes, concerns, and information sources reported by parents of young children: results from the 2009 HealthStyles survey. *Pediatrics* 2011; **127**(Suppl 1): S92–S99.

417 Gust DA, Campbell S, Kennedy A, Shui I, Barker L, Schwartz B. Parental concerns and medical-seeking behavior after immunization. *Am J Prev Med* 2006; **31**(1): 32–5.

418 Salmon DA, Sotir MJ, Pan WK, *et al*. Parental vaccine refusal in Wisconsin: a case–control study. *WMJ* 2009; **108**(1): 17–23.

419 Robison SG, Groom H, Young C. Frequency of alternative immunization schedule use in a metropolitan area. *Pediatrics* 2012; **130**(1): 32–8.

420 Institute of Medicine. *The Childhood Immunization Schedule and Safety: Stakeholder Concerns, Scientific Evidence, and Future Studies*. Washington, DC: Institute of Medicine, 2013.

421 Glanz JM, McClure DL, Magid DJ, *et al*. Parental refusal of pertussis vaccination is associated with an increased risk of pertussis infection in children. *Pediatrics* 2009; **123**(6): 1446–51.

422 Opel DJ, Taylor JA, Zhou C, Catz S, Myaing M, Mangione-Smith R. The relationship between parent attitudes about childhood vaccines survey scores and future child immunization status: a validation study. *JAMA Pediatr* 2013; **167**(11): 1065–71.

423 Luman ET, Barker LE, Shaw KM, McCauley MM, Buehler JW, Pickering LK. Timeliness of childhood vaccinations in the United States: days undervaccinated and number of vaccines delayed. *JAMA* 2005; **293**(10): 1204–11.

424 Opel DJ, Taylor JA, Mangione-Smith R, *et al*. Validity and reliability of a survey to identify vaccine-hesitant parents. *Vaccine* 2011; **29**(38): 6598–605.

425 Daley MF, Glanz JM, Newcomer SR, *et al*. Assessing misclassification of vaccination status: implications for studies of the safety of the childhood immunization schedule. *Vaccine* 2017; **35**(15): 1873–8.

426 Iskander J, Pool V, Zhou W, English-Bullard R. Data mining in the US using the Vaccine Adverse Event Reporting System. *Drug Saf* 2006; **29**(5): 375–84.

427 Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. *Biometrics* 2003; **59**(2): 323–31.

428 Platt R, Madre L, Reynolds R, Tilson H. Active drug safety surveillance: a tool to improve public health. *Pharmacoepidemiol Drug Saf* 2008; **17**(12): 1175–82.

429 Onions D, Cote C, Love B, *et al*. Ensuring the safety of vaccine cell substrates by massively parallel sequencing of the transcriptome. *Vaccine* 2011; **29**(41): 7117–21.

430 Sahoo MK, Holubar M, Huang C, *et al*. Detection of emerging vaccine-related polioviruses by deep sequencing. *J Clin Microbiol* 2017; **55**(7): 2162–71.

431 Prachi P, Donati C, Masciopinto F, Rappuoli R, Bagnoli F. Deep sequencing in pre- and clinical vaccine research. *Public Health Genomics* 2013; **16**(1–2): 62–8.

432 Tsai FJ, Lee H, Fan VY. Perspective and investments in health system strengthening of Gavi, the Vaccine Alliance: a content analysis of health system strengthening-specific funding. *Int Health* 2016; **8**(4): 246–52.

21

Epidemiologic Studies of Medical Devices: Methodologic Considerations for Implantable Devices

Danica Marinac-Dabic[1], Sharon-Lise Normand[2], Art Sedrakyan[3], and Thomas P. Gross[4]

[1] *Office of Clinical Evidence and Analysis, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA*
[2] *Department of Health Care Policy, Harvard Medical School and Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA*
[3] *Weill Cornell Medical College, New York, NY, USA*
[4] *Office of Surveillance and Biometrics, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA*

Medical devices save lives, prevent disease, improve health, and help patients live productive lives. The last decade has seen an explosion in medical device technologies worldwide. It has been recently estimated that the global medical device market exceeded $389 billion in 2017 [1].

Groundbreaking innovations in the areas of transcatheter interventions, nanotechnology, telemedicine, robotic procedures, sophisticated health information technology software, and smart applications continue to offer new diagnostic and therapeutic options to patients and clinicians. Recent approval of the first medical device that uses artificial intelligence to aid in diagnosis of retinopathy is a good example of the dynamic landscape where software as a medical device will play a more prominent role in the delivery of healthcare [2].

From the public health perspective, the availability of diverse health technologies demands both the capacity and commitment to development of new methodologic approaches for evidence generation, appraisal, and synthesis. Medical device epidemiology is well suited not only to studying the extent of utilization of medical devices, but also to studying utilization patterns and quantifying risk/benefit for certain outcomes in defined populations. Furthermore, an attractive feature of modern device epidemiology is the implementation of techniques to integrate information available from the growing body of heterogeneous data. Operating at the intersection of scientific knowledge and healthcare, it is the practice of epidemiology that ensures consistently reliable approaches to combine and update information in order to maximize quality, minimize bias, and reduce uncertainty in understanding the risks and benefits of new devices.

What Is a Medical Device and How Is It Different from a Drug?

The definition of a medical device varies somewhat by country (Table 21.1). The US government defines a medical device as:

> an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent or other similar or related article, including any component part or accessory which is: (1) recognized in the official National Formulary, or United

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

States Pharmacopeia or any supplement of them; (2) intended for use in the diagnosis of disease or other conditions or in the cure, mitigation, treatment or prevention of disease in men or other animals; or (3) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its principal intended purposes. [3]

Definitions used in the European Union (EU), Canada, Australia, and Japan are slightly different (Table 21.1) [3–9].

There is a long history of international harmonization efforts in the regulation of devices. The Global Harmonization Task Force (GHTF), formed in 1992 in an effort to respond to the growing need for international harmonization in the regulation of medical devices, produced a harmonized definition of a medical device [9]. To build on the strong foundation of the GHTF, the International Medical Device Regulators Forum (IMDRF) emerged in 2011 as a driver of harmonization and convergence efforts among regulatory authorities on medical devices [10]. This voluntary organization brings together regulators from Australia, the EU, Canada, the US, Singapore, China, Japan, and Brazil. The IMDRF has made major steps toward a

Table 21.1 International definitions of medical device.

Country	Definition of Medical Device
United States	An instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent or other similar or related article, including any component part or accessory which is: 1) recognized in the official National Formulary, or United States Pharmacopeia or any supplement of them; 2) intended for use in the diagnosis of disease or other conditions or in the cure, mitigation, treatment or prevention of disease in men or other animals, or, 3) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its principal intended purposes.
European Union	"Medical Device" means any instrument, apparatus, appliance, material or other article, whether used alone or in combination, including the software necessary for its proper application intended by the manufacturer to be used for human beings for the purpose of: • diagnosis, prevention, monitoring, treatment or alleviation of disease, • diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap, investigation, replacement or modification of the anatomy or of a physiological process, control of conception, and which does not achieve its principal intended action in or on the human body by pharmacologic, immunologic or metabolic means, but which may be assisted in its function by such means. "Accessory" means an article which whilst not being a device is intended specifically by its manufacturer to be used together with a device to enable it to be used in accordance with the use of the device intended by the manufacturer of the device. "Device used for in vitro diagnosis" means any device which is a reagent, reagent product, kit, instrument, equipment or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of samples derived from the human body with a view to providing information on the physiological state, state of health or disease, or congenital abnormality thereof.

(*Continued*)

Table 21.1 (Continued)

Country	Definition of Medical Device
Canada	Any article, instrument, apparatus, or contrivance, including a component, part or accessory thereof, manufactured, sold or represented for use in a) the diagnosis, treatment, mitigation or prevention of a disease, disorder or abnormal physical state, or its symptoms, in human beings or animals, b) restoring, correcting or modifying a body function or the body structure of human beings or animals, c) the diagnosis of pregnancy in human beings or animals, or d) the care of human beings or animals during pregnancy and at and after birth of the offspring, and includes a contraceptive device but does not include a drug.
Australia	Any instrument, apparatus, appliance, material or other article (whether used alone or in combination, and including the software necessary for its proper application) intended by the person under whose name it is to be supplied, to be used for human beings for the purposes of one or more of the following: • diagnosis, prevention, monitoring, treatment or alleviation of disease, • diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap, • investigation, replacement or modification of the anatomy or of a physiological process, • control of conception, and does not achieve its principal intended action in or on the human body by pharmacologic, immunologic or metabolic means, but which may be assisted in its function by such means; or an accessory to such an instrument, apparatus, appliance, material or other article.
Japan	Instruments and apparatus which are intended for use in the diagnosis, cure or prevention of diseases in man or animals, or intended to affect the structure or any function of the body of man or other animals, and which are designated by Cabinet Order.

convergence of activities in the area of adverse event reporting, patient registries, software as a medical device, and the implementation of Unique Device Identification (UDI), among other efforts [11,12].

While there are similarities among countries regarding medical device definitions and classifications, the differences exist in the rigor required for device approval. For example, before a medical device is allowed to enter the US market, a reasonable assurance of its safety and effectiveness must be established. Operationally, this goal is accomplished through the US Food and Drug Administration's (FDA) use of regulatory controls and the classification process. Based on the varying levels of potential benefit and risk, devices are classified into one of three regulatory classes. Class I devices (e.g., elastic bandages, surgical gloves, manual surgical instruments) are subject to general controls such as labeling and good manufacturing processes. These devices present minimal potential for harm to the patient and require neither clinical testing nor special controls to establish a reasonable assurance of safety and effectiveness. Class II devices by definition are higher risk (e.g., infusion pumps, diagnostic ultrasound machines) and are subject not only to general controls, but also to special controls such as guidance and standards. Certain of these devices may require additional clinical testing as well. Medical devices with the highest level of risk (e.g., implantable deep brain stimulators, coronary stents, and hip resurfacing systems) are categorized as class III and receive the highest level of scrutiny for regulatory

approval. In addition to general and special controls, these devices require submission of clinical data in support of premarket submissions. In addition, the effectiveness and safety of these devices need to be determined based on valid scientific evidence, defined as "evidence from well controlled investigations, partially controlled studies, studies and objective trials without matched controls, well documented case histories, by qualified experts, and reports of significant human experience from a marketed device. More recent regulatory initiatives introduced a number of new opportunities for shifting scientific evidence generation to the postmarket setting and leveraging real-world evidence for regulatory decision-making" [13–16].

Other countries have similar classifications. In Canada, devices of classes III and IV are subject to in-depth regulatory evaluation, while class II devices require only the manufacturer's declaration of device safety and effectiveness, and class I devices are exempted from premarket submission. In the EU, manufacturers of devices of classes II and III, as well as devices of class I with either a measuring function or sterility requirements, must submit to the regulator (competent authority) (i) a Declaration of Conformity to the appropriate European Commission (EC) Directives, and (ii) details of the conformity assessment procedure followed. In addition, for higher-risk class devices that require design examination or type examination, the corresponding EC Certificates issued by a notified body must also be submitted to the competent authority [17]. In Australia, the risk-based evaluation system also requires that all "registrable" devices must undergo rigorous premarket evaluation before market entry; "listable" devices are less rigorously regulated, but may be evaluated for safety (not efficacy) if there are regulatory concerns about the risk profile of the product. Registrable products include prescription and nonprescription medicines and implantable medical devices (e.g., active implantables and devices of animal origin), while listable goods include vitamins, minerals, herbal medicines, sunscreens, and most medical devices [18]. In Japan, all devices above class II must obtain a central government license for market entry.

Many countries are in the process of implementing new regulatory requirements to strengthen regulatory evidence, while relying more on the data generated by patient registries and other real-world data sources. The common regulatory theme underpinning classification and requirements across regions and countries is the level of risk, despite the differences in how the risk is defined.

In this chapter, we concentrate on implantable devices (US classes II and III) because of their significant public health impact, high risk for adverse events, and uncertainties surrounding the effects of long-term exposure.

Implantable medical devices comprise an important device category in the very heterogeneous world of medical devices. As is true of other devices, implantables share characteristics that distinguish them not only from other devices, but also from regulated drugs. Table 21.2 highlights the characteristics that are further distinguished between medical devices or drugs in general.

Implantable devices, most of which are in class III (highest risk), can be further differentiated. This device category in general has longer product life cycles, although incremental changes do occur. Implantable devices may comprise multiple components (such as a total hip implant) or single components (such as a pacemaker lead). Exposures to such devices are typically chronic (an exception being temporary implantables, such as inferior vena cava filters), with the onset of exposure clearly defined at the time of implantation. Exposure ends at the time of device removal, but may not be clear cut if part of the device remains (e.g., in case of silicone leakage from ruptured breast implants, or components that are absorbed over time such as with biologically based dermal fillers or barrier adhesion devices).

Table 21.2 Characteristics of medical devices as compared to drugs (US regulations).

	Characteristics of medical devices as compared to drugs	
Characteristic	**Device**	**Drug**
Product life cycles	Short to long	Short
Incremental changes	Common	Rare
Equivalence	Technological (class I and II)	Therapeutic (for generics)
Clinical trials required	n = 1 (class III, some II)	n = 2 (NDAs)
Trial reimbursement	Frequent	Rare
Orphan designation	4000	200 000
Assuring manufacturing quality	ISO 9000	Good Manufacturing Practices (cGMP)
Required postmarket studies	Postapproval studies Section 522 studies	Phase IV studies
Product identification	Product codes Unique Device Identifiers	NDCs
Components/ingredients	Single or multiple (may change over time)	Single or multiple
Exposures	Acute, chronic, intermittent, episodic	Similar
Stopping exposure	Simple to complex	Typically, simple
User interface	Patient or clinician	Typically, patient
Users of same product	Single or multiple	Single
How used	Single use/disposable, reusable, implantable, durable	Typically, multiple use
Product effects	Typically localized Acute or chronic	Systemic, Typically, acute
Product hazards	Human factors	Less prominent
	Noncompliance	Same
	Interactions	More prominent
	Malfunctions (manufacturing)	Drug quality problems
	Environmental hazards	Same
	Toxic/allergic	More prominent
	Packaging defects	Same
	Software glitches	
	Poor maintenance	

NDA, New Drug Application; NDC, National Drug Code.

Outcomes associated with implantable devices are affected not only by underlying patient factors and device factors (such as biomaterials), but also, importantly, by user interface (e.g., operator technique, operator experience). Adverse effects of implantable devices are typically localized, but may be more systemic (e.g., secondary to toxic, inflammatory, allergic, autoimmune effects). Additional hazards may be related to human factors (e.g., proper programming of pacemakers) and interactions (e.g., magnetic resonance imaging [MRI] interaction with deep brain stimulator leads). Lastly, malfunctions may derive from several sources, including manufacturing problems, design-induced errors, and anatomic or engineering effects (e.g., repetitive flexing of an implantable cardioverter defibrillator lead causing fracture).

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Diffusion to Clinical Practice and Utilization

The introduction and adoption of new medical devices may induce a breakthrough transformation of clinical practice [19]. The rate of diffusion of medical device technology to clinical practice is influenced by many factors, including device complexity, relative advantage when compared with similar available treatments, positions of opinion-leading organizations, healthcare reimbursement decisions, commercial competition, evidence-based guidelines, publication of key research papers, regulatory actions, medical liability issues, and legislative environment [20]. Accordingly, the adoption of different new devices and their dissemination into routine clinical practice often follow different and sometimes unpredictable patterns. These diffusion pathways may contribute considerably to variations in patient outcomes and may have a prominent impact on the frequency of adverse events seen in the usual healthcare settings, when compared to patient experience in premarket clinical trials. An epidemiologic assessment of factors that can potentially influence the adoption of new technology can inform regulatory science (both premarket and postmarket), help in the development of clinical guidelines and policies, and can significantly shape national reimbursement strategies [21].

Evaluating Device Performance in the Real World

When traditional randomized clinical trials (RCTs) for new devices are planned, elite investigators are often selected to participate. For example, in RCTs of devices used in surgery, the operators are typically early adopters, highly skilled, and quick learners, which influences the learning curve. The learning curve can be defined as a constant proportional improvement in performance, such as clinical outcomes of medical procedure, with each doubling of cumulative experience [22]. The rate of learning (the shape of the learning curve) and the interaction with other variables for an average surgeon are very difficult to gauge from RCTs, which include only a small number of selected surgeons. It is only once the approved device enters clinical practice where observational data are collected on patient outcomes for a wide range of surgeons and hospitals that it is possible to estimate the effect of the learning curve on patient outcomes.

The operator's learning curve can be steep, protracted, or anywhere in between, and can have a substantial impact on patient outcomes. Traditionally, the learning curve is studied using the volume–outcome relationship. Some volume–outcome studies have demonstrated that increased surgical volume has an inverse relationship with the likelihood of poor outcomes such as complications, revision surgery, length of stay, and mortality

[23–25]. Other studies have shown a volume threshold for procedures, above which increasing volume is no longer associated with improved outcomes [26,27].

Lastly, other researchers have noted a trimodal institutional learning curve: rapid initial phase, followed by declining success, representing new adopters, and then recovery to an improved steady state [28]. These observations indicate three distinct components of the volume–outcome relationship that can be studied: (i) lifetime experience (operator's volume); (ii) operator's annual volume; and (iii) hospital volume where operators practice. Other factors beyond volume that relate to the learning curve, including the type of procedure (e.g., diagnostic vs. interventional) and practice setting (e.g., institutional teaching status), are also readily available in many observational databases and can be studied. Adequate study of learning curves can establish thresholds for proficiency based on background expertise related to physicians' specialties [29]. This has been the case with the stenting of carotid arteries by operators from varying specialties (e.g., radiologists, cardiologists, and neurosurgeons).

By the nature of their design, RCTs involve select, nonrepresentative patient populations. A number of studies have highlighted the disparities in disease prevalence, progression, and health outcomes of medical device technology in subgroups of the population [30,31]. Premarket medical device trials often lack sufficient representation of important patient populations (women, children, the elderly, racial and ethnic minorities, and others), which hampers the application of the results to real-world populations. Well-designed observational studies can provide more information on device performance in the subpopulations of interest in a real-world setting. The public health utility of observational studies has been increasing, with advances in medical device data capture in medical records, electronic databases, and prospective registries, and the development of innovative analytic tools using observational data [32,33].

The recent increase in US interest in comparative effectiveness, facilitated by the American Recovery and Reinvestment Act of 2009 and subsequent healthcare reform legislation [34], has begun focusing the national attention on building methods and infrastructure for emerging comparative effectiveness research (CER; see Chapter 26). CER, as defined by the US Institute of Medicine (IOM), is "generation and synthesis of evidence that compare the benefits and harms of alternative methods to prevent, diagnose, treat and monitor or improve the delivery of care" to "assist consumers, clinicians, purchasers and policy makers to make informed decisions that will improve health care at both the individual and population levels." CER is vital for the improvement of healthcare quality, better regulatory decisions, and thoughtful guidelines for clinicians and patients [35].

In recent years, several other countries have established agencies to evaluate health technologies and inform healthcare policy decisions [36]. These organizations are different in terms of structures, methods, and processes, but in all of them CER is an effort that aims to address the needs of payers, patients, clinical professionals, and policymakers. Rapid growth of new medical devices, modifications of existing models, and dramatically shorter device life cycles will continue to create demands for dynamic and up-to-date comparative effectiveness and safety efforts [37,38]. Epidemiologic research will play a prominent role in medical device evidence synthesis in the context of health technology evaluation

Long-Term Safety and Effectiveness

Device premarket clinical trials are typically of short duration (e.g., 1–2 years) and generate limited information on long-term safety and effectiveness. Due to the inherent complexity of

implantable devices, it is often difficult to predict fully their long-term safety and effectiveness based solely on preclinical testing and premarket clinical trials. The FDA's postmarket attention is therefore increasingly directed toward ensuring that studies/surveillance of sufficient size and length of follow-up are conducted in the postmarket setting to better illustrate and assess problems occurring long term [39]. Other countries have established national registries of procedures involving implanted medical devices that collect long-term patient outcomes and device performance (e.g., orthopedic registries in Sweden, the UK, Australia, Canada, and other countries) [40–43]. During the past decades the US has also seen a growth in national registries (e.g., National Cardiovascular Data Registry, Kaiser Permanente National Joint Replacement Registry (NJRR), American Joint Replacement Registry (AJRR), and National Breast Implant Registry (NBIR) [44].

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

Evidence generation for implantable medical devices requires accounting for unique issues that typically do not arise when evaluating the benefits and risks of drugs. Key issues arise from the interaction of device, operator, and the setting (e.g., hospital, outpatient clinic) in which the device is being used. Furthermore, the device's design, complexity, and specific biomaterial and mechanical characteristics can be as important to outcomes as the device's clinical applications, such as the type of the lesion being treated, the severity of the disease, and concomitant therapy. In the commonly used device epidemiologic research databases, these details are often only partially available, and sometimes are missing.

Challenges in Individual Patient Exposure Assessment

The UDI captures critical device information, such as manufacturer name, brand, version or model, and device group terms. The UDI system has only recently been established, and the UDI is now entering data systems within the US, unlike pharmaceuticals, where the National Drug Code (NDC) Directory has a long history and is broadly used. Consequently, many databases still lack specific device identifiers, making exposure assessments challenging. For example, procedure codes may capture device groups (such as hip implants), but lack specificity to the manufacturer level. Characterization of the sensitivity and specificity of device identifiers found in medical records, clinical registries, or insurance claims databases will be important to understand errors in device exposure. Promoting routine documentation of UDIs in medical records and in other health databases will contribute to a better understanding of device-specific performance [45].

Another challenge associated with medical device epidemiology relates to device complexity: devices are frequently approved or used as *systems* involving several components. Device components are often used in combination with components of the same or different brands. Thus, experience with capturing complete device exposure information is far more complex for devices than it is for drugs. Once completely adopted, such a robust, widely incorporated medical device nomenclature will significantly further safety surveillance and epidemiologic studies of medical devices [46].

Challenges in National Population Exposure Assessment

Incident and prevalent exposure data provide the necessary context for interpretation of the possible relationship between device exposure and outcome. Therefore, strategies to develop

the necessary infrastructure, methodologies, and partnerships will have to include incorporation of UDIs into data systems, including electronic health records, and routine documentation of device use and patient problems associated with that use. Since 2010, under its Medical Device Epidemiology Network (MDEpiNet) Initiative, the FDA has worked with its partners to launch a series of strategic demonstration projects designed to advance and test-drive novel approaches to methods and infrastructure development. To build on this foundational work, in 2012 the FDA launched a strategy to work with all stakeholders to develop a national system for evaluation of medical devices [26–28], leading to the establishment of the National Evaluation System for health Technology (NEST) Coordinating Center and partnering organizations [47].

Currently, in a less-than-ideal national surveillance environment, population exposure data for medical devices have to be derived from a variety of sources including, among others, medical billing claims data, registries, national surveys, nationally representative samples of providers, and market data. In addition, these sources differ in their level of device specificity (e.g., by device group in claims data compared to specific brand in registries). While these sources differ in the level of completeness and reliability, they may complement each other.

Challenges in Comparative Studies

Epidemiologic population-based research relies on nonexperimental (or observational) approaches to develop evidence about the safety and effectiveness of medical products. While there is a recognition of the limitations related to observational data as compared to RCTs [48], we also need to recognize two facts. First, trial design characteristics such as randomization, allocation concealment, and masking/blinding, which are recognized as critical characteristics of high-quality studies, are often difficult to adopt in device trials. Second, observational data frequently complement experimental data rather than replace them. In addition, observational data derived from routine clinical practice are required to determine if the device can be successfully deployed or implanted (learning curve issues) outside of centers and subjects who participate in clinical trials, to quantify risks of adverse events in larger populations, and to assess any modifications to the product. These features are very real issues that affect the safety and effectiveness of medical devices, but are only partially applicable to pharmaceuticals.

Ensuring Comparability of Study Groups

Cohort designs offer the opportunity to create comparable groups of patients exposed to devices of interest. These observational studies should take advantage of statistical adjustments for known and measured confounders, rely on models with few parametric assumptions, and employ methods that demonstrate robustness of findings to deviations from assumptions [4]. For example [49], the "no unmeasured confounder" assumption underlies most epidemiologic studies, and methods are available to characterize the impact of residual confounding (see also Chapter 43). Prospective data collection can minimize the risk of unmeasured confounders through *a priori* determination and collection of such factors. Consecutive patient enrollment provides critical information to inform operator learning behaviors.

We have good tools to address unequal distribution in observed patient characteristics (confounders) that is not severely confounded by indication (see Chapter 43). Most of the adjustment techniques deal with imbalances in confounding factors between the study groups. In addition to addressing known imbalances, one can also theoretically remove the bias related to unobserved prognostic factors if the unobserved factors are highly correlated with the measured prognostic factors [50]. Of course,

this is an assumption which cannot be tested, but the size of the unmeasured confounders required to change the results can be quantified and reported. Additionally, including a "falsification outcome" may help diagnose residual confounding. For example, implantation of a coronary stent should not reduce in-hospital mortality.

Several analytic methods are available to deal with selection factors and confounding. These methods involve stratification, regression models, or a combination of the two using a propensity score [51,52]. Each approach relies on a set of assumptions, which may or may not be appropriate in the particular setting. When it is felt that there is unmeasured confounding present beyond that accounted for in the collected information, another potential approach is using instrumental variable-based methods, if a valid instrument can be found (see Chapter 43) [53–55].

Addressing Issues of Sample Size, Real-World Performance, and Long-Term Outcomes

As mentioned previously, premarket clinical trials are powered first and foremost for efficacy outcomes. Powering these trials based on less common or rare but serious side effects is not feasible in most instances (see Chapter 4). As is true for drugs, clinical trials of devices, because of small sample size and participant selection, often lack generalizability, defined as the extension of research findings and conclusions from a study conducted on a target population to the population at large.

Systematic reviews with meta-analyses (see Chapter 36) are observational studies that attempt to capitalize on the detailed data collection within the studies that are the subject of the review. Systematic reviews are one mechanism to address the small study problems of clinical trials. Systematic reviews with meta-analysis are based on the premise that most of the individual clinical trials of devices and surgery carefully record relevant clinical outcomes and offer a great opportunity to conduct evidence appraisal and synthesis when a reasonable number of studies are available.

Well-designed cohort studies are often large and involve consecutive patient enrollment and data collection that is comprehensive. They are the best-suited tools to evaluate the safety and effectiveness of devices in the real-world populations and are based on solid scientific knowledge accumulated about the devices.

With the explosion of electronic data acquisition and exchange technologies, computing resources, and linkability across data sources, newer statistical methods that exploit high-dimensional data (e.g., the number of features can exceed the number of observations) provide more opportunities for medical device epidemiology researchers. With higher-dimensional data, additional assumptions are required to prevent overfitting or to ensure that a unique solution exists. These additional assumptions are implemented using "regularization methods" that effectively penalize the inclusion of many predictors. One group of regularization methods assumes sparseness, that a small number of variables represent the underlying data structure. Approaches such as LASSO (least absolute shrinkage and selection operator) [56], sparse additive models [57], or sparse prior distributions for regression coefficients are examples in this group. For instance, a weakly informative default prior distribution for regression coefficients that centers the coefficients at zero but permits nonzero coefficients through the use of heavy tails falls into this category [58]. Another group of methods combines a large number of models in order to obtain a stronger model by averaging across models. Bagging (e.g., random forests), boosting, and stacking (e.g., Super Learner) [59,60] are algorithms that fall into this category. The intuition for these methods exploits the belief that more than one model may "fit" the data, so that averaging across multiple models may lead to better conclusions. With large observational studies, one

can evaluate relevant subgroup effects as well as rare safety and effectiveness endpoints that cannot be captured by RCTs.

Currently Available Solutions

Passive Surveillance

Once a device has received marketing approval, manufacturers must follow Good Manufacturing Practices and monitor the safety of their products, including keeping a complaint file and forwarding reports of adverse events to the regulatory authorities. For example, in the US manufacturers are required to submit reports of device-related deaths, serious injuries, and malfunctions to the FDA. Healthcare providers and consumers submit reports voluntarily through MedWatch [61]. These reports, obtained through passive surveillance (see also Chapter 10), are housed in the Manufacturer and User Facility Device Experience (MAUDE) database, established in 1996. As of May 2018, MAUDE contained more than 6 million reports and received approximately 80 000–110 000 reports per month. The vast majority of reports are from manufacturers, with a small percentage from user facilities, voluntary sources, and importers [62].

Regulatory agencies, including those participating in IMDRF, have the authority to monitor the safety of devices by reviewing adverse event reports from users, sponsors, other available data sources, and the scientific literature [9]. In assessing these reports, in addition to specific patient characteristics, regulatory bodies consider the following factors: failure potential resulting from design or manufacturing problems; use error potential from improper device assembly, misreading instructions, or improper surgical technique; incorrect clinical use; and inadequate instructions for use. Possible packaging errors, support system failure, adverse environmental factors, maintenance error, adverse device interactions such as electromagnetic interference, or toxic/idiosyncratic reactions are also considered [63]. Some manufacturers conduct failure analyses on retained or returned products (including implantables) in the event of a reported device problem.

To enhance the usefulness of reported data, statistical tools are used to assist in detecting new signals [64,65] (see Chapter 27). Bayesian and other data-mining methods are used to estimate the relative frequency of specific adverse event–device combinations, as compared to the frequency of the event with all other devices (in the same group) in the database. To aid in this effort, and reporting and signal detection in general, an extensive hierarchical vocabulary for adverse device outcomes (e.g., high impedance in pacemakers) also has been developed [66].

Passive reporting systems have noticeable weaknesses, including (i) data may be incomplete or inaccurate and are typically not independently verified; (ii) data may reflect reporting biases driven by event severity or uniqueness or publicity and litigation; (iii) causality cannot be inferred from any individual report (see Chapter 29); and (iv) events are generally underreported and this, in combination with a lack of denominator (exposure) data, precludes the determination of event incidence or prevalence. The latter point is particularly important for implantable devices, since reports may capture device-associated events (such as thrombosis, infection, stroke, revision, or replacement) for which estimation of incidence is of paramount importance.

Enhanced Surveillance

To enhance the understanding of clinical issues for medical devices, the Medical Product Safety Network (MedSun) was established to provide national medical device surveillance based on a subset of user facilities in the US [67,68]. MedSun currently includes approximately 350 hospitals nationwide. Through its ongoing bidirectional interactions, including educational

fora, problem-solving and posting of reports, and targeted surveys, this enhanced surveillance network provides more timely amplification of potential safety signals.

Reports received through passive and enhanced systems have resulted in significant public health notifications, including those related to injuries (i) from transvaginal placement of surgical mesh [69]; (ii) associated with use of recombinant bone morphogenetic protein in cervical spine fusion [70]; and (iii) from MRI-induced interactions in patients with implanted neurologic stimulators [71]. Reports received have also spurred the development of significant observational studies elucidating risk factors for meningitis associated with cochlear implants [72], and hemorrhagic complications associated with the use of hemostasis devices, including one high-risk device [73,74].

Signal Detection/Outlier Identification Using a Variety of Data Sources and Methods

Cumulative Sum of Outcomes (CUSUM) Methodology

CUSUM is a sequential statistical analysis method with graphic application that allows online identification of changing rates of device failure or surgical complications. For example, a likelihood-based scoring method of calculation of CUSUM is used by the Scottish Orthopedic registry, described as part of the International Consortium of Orthopedics (ICOR) series [75]. Outlier device/surgeon status is identified at a point set in advance and is named the prediction limit. Setting the statistical thresholds at agreed levels helps balance the risk of failure against that of false alerts. Setting a prediction limit is not an exact science, and changing the statistical criteria will change outlier identification. Hence, the results should be interpreted as a potential signal that does not yet mean a poorly performing implant or device in general. One of the advantages of the CUSUM method is the ability to track both surgeons and the introduction of the new device to evaluate the surgeon/surgical team/device "combination." For example, CUSUM allows tracking of outcomes of high-volume surgeons with changes in practice over time, and determination of periods of outlier performance that were linked with the introduction of new implants [76].

Cumulative Revision Rate over Time

Depicting an unadjusted cumulative revision rate over time after implantation of the device is a simple but powerful technique allowing identification of outlier implants when compared to the overall or group average. The method also allows calculation of accompanying 95% confidence intervals using various methods. For example, the Australian orthopedic registry process identified the ASR artificial hip as an outlier device using this method followed by Cox proportional-hazards modeling to calculate the hazard ratios and adjust for age and sex, in order to conduct a comparative analysis of revision rate between groups [77].

Funnel Plots

Another graphic approach is funnel plots, which are based on the application of Shewhart charts in medicine [78]. Through the use of funnel plots such as that shown in Figure 21.1, it is possible to compare the observed events (e.g., specific device failure) against the national average within the population [79]. For instance, devices falling above the 95% or 99.8% control limits (set in advance) for risk are deemed outliers. For various true event rates around the gold-standard rate, the funnel plot shows which devices can be called outliers. Like several other methods, this approach is heavily dependent on assumptions about equivalent underlying risk. If there is heterogeneity in the underlying risk (as might occur with differing standards of care across sites, differing expertise of operators, or differing disease progression among patients between sites), then departures outside the limits may be more

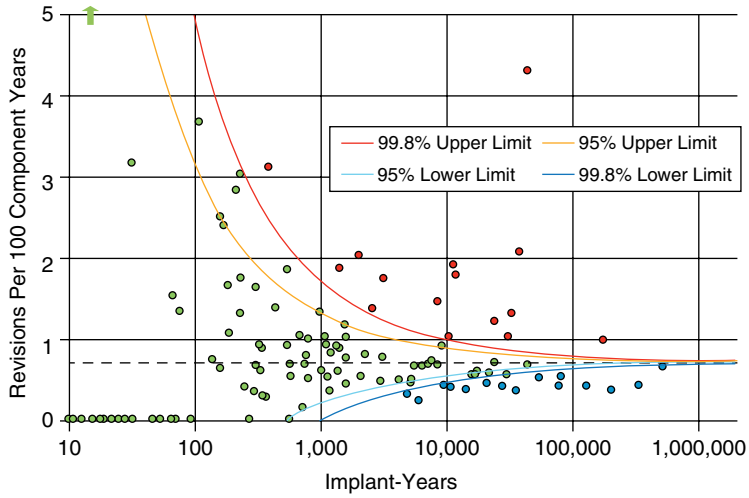


Figure 21.1 Hypothetical funnel plot. In this hypothetical funnel plot from the Joint Replacement Registry, each circle represents one device. The X axis denotes the number of devices combined with the number of years followed up for a particular device tracked by the registry. The Y axis represents the "true" event rate (unobserved). Devices falling above 95% or 99.8% control limits (set in advance) for risk are deemed outliers. For various "true" event rates around the gold-standard rate, the funnel plot shows which devices can be called outliers.

reflective of issues with the assumptions broadly, than with issues in the performance of those points outside the limits [12].

Automated Surveillance

In-hospital safety signals (such as myocardial infarction) for recently introduced interventional cardiovascular devices (such as drug-eluting coronary stents) are possible to explore using an automated computerized safety surveillance system, DELTA (Data Extraction and Longitudinal Trend Analysis), which was designed to support flexible safety surveillance applicable to a broad range of medical devices [80,81]. The system is compatible with a broad array of potential data sources and supports a variety of statistical methods, allowing for both unadjusted and risk-adjusted safety monitoring for prospective and retrospective analyses.

Data Linkages for Long-Term Follow-Up

Linking registry data to other data sources is often necessary to capture longitudinal data [82] and yield an enriched data source for regulatory purposes, clinical decision-making, and reimbursement purposes. In addition, linkage is often used for validation processes of patient registries. Complementary data may include but are not limited to other registries, national death records, electronic medical records, or longitudinal administrative medical claims/discharge databases.

There are two broad methods of data linkage:

- ***Deterministic (direct)***. Deterministic linkage algorithms aim to determine if record pairs agree or disagree on the available set of identifiers and when there is agreement on a given identifier.
- ***Probabilistic (indirect)***. Probabilistic approaches to link large datasets aim to use limited identifiers applied methodologically in a way that maximizes the probability that a data field agrees given that a record pair matches, minimizes the probability that a data field agrees given that a record pair is unmatched, and provides greater precision from nonuniformly distributed fields [83]. It is a method that enables the combination of record information in different datasets to form a new linked dataset. The probabilistic link uses several identifiers, in combination, to identify and evaluate links. Probabilistic binding is usually used when a unique identifier is not available or is of insufficient quality

(Australian Government, National Statistical Service, Data Linking) [12].

Both matching approaches have their strengths and their limitations. It is generally recommended to evaluate the probability of successful matching as a rule, and then employ a combination of deterministic and probabilistic methods that optimizes the combination of completeness of the population and accuracy of matching [12].

Registries and Strategically Coordinated Registry Networks

Patient registries (see also Chapter 16) constitute a great infrastructure for conducting large-scale medical device studies. A pragmatic and adequate medical device registry is defined in the International Medical Device Regulators Forum (IMDRF) report [11] and the subsequent *Lancet* publications [84] as an

> Organized system with a primary aim to improve the quality of patient care that continuously collects relevant data, evaluates meaningful outcomes and comprehensively covers the population defined by exposure to particular device(s) at a reasonably generalizable scale (e.g. international, national, regional, and health system).

One important value that registries add is their ability to collect data on large numbers of patients and clinical settings and, through linkages (see earlier discussion), on outcomes over time. Device registries have been historically designed to address issues related to clinical performance and quality improvements in healthcare. They have been more recently involved in comparative research, assessing technology for reimbursement, and monitoring postmarket device performance. Depending on the questions to be addressed, registries can be designed to capture data on exposures and/or conditions of interest, type of healthcare service delivered (e.g., surgical treatment or diagnostic procedure), or outcome of interest (e.g., adverse event, disorder, or disease). In the absence of a UDI code in many healthcare data sources, the added value of registries includes capturing brand/model-specific information crucial for signal identification and comparative effectiveness/safety studies. The complexity and scientific rigor of a registry can vary from those designed to evaluate the quality of healthcare delivered, those specifically established to study the sustained effectiveness and safety of a specific procedure, and those designed to systematically collect long-term data on many different types of treatment, including risk factors, clinical events, and outcomes in a defined population. Once the framework of a registry is in place, studies with various designs can be performed using registry data (cohort, case–cohort, case–control, cross-sectional, quasi-experimental), both mandated and discretionary.

To build on the concept put forward in the 2015 MDEpiNet National Medical Device Registry Task Force Report, the medical device ecosystem (including industry, regulators, the clinical community, public health agencies, and academia) increased efforts to foster the development of strategically coordinated registry networks (CRNs) as a valuable tool for capturing the utilization of devices, identifying early signals, and studying the postmarket performance of medical technology. The CRN is an analytic paradigm with an emphasis on the development of strategically partnered electronic health information systems that support both the implementation of structured device identifiers, core minimum data elements, and definitions within registries, and the ability to share complementary data across a number of information systems [85].

In the past seven years, the MDEpiNet Science and Infrastructure Center has implemented more than 14 CRNs (see the examples in Table 21.3). The existing CRNs are used to monitor the performance of high-risk devices,

Table 21.3 Selected examples of strategically coordinated registry networks (CRN) in the US and their international registry consortia.

CRN Name – US	Examples of medical devices captured	US registry members	Linkage to other data sources	International registry consortium
Vascular Implants and Interventions Surveillance and Outcomes Network (VISION)	Carotid artery stents Abdominal aortic aneurysm (AAA) Peripheral stents	Vascular Quality Initiative (VQI)	Administrative claims data	International Consortium of Vascular Registries (ICVR)
Orthopedic Devices CRN (Ortho-CRN)	Hip and knee arthroplasty devices	AJRR – American Joint Replacement Registry Kaiser Permanente NJRR – National Joint Replacement Registry MARCQI – Michigan Arthroplasty Registry Collaborative Quality Initiative Health East Arthroplasty Registry Virginia Arthroplasty Registry FORCE TJR – Function and Outcome Research for Comparative Effectiveness in Total Joint Replacement Registry	Administrative claims data	International Consortium of Orthopedic Registries (ICOR)
National Breast Implants Registry (NBIR) CRN	Breast Implants (silicone, saline)	NBIR – National Breast Implants Registry PROFILE – Registry for Breast Implants and Anaplastic Large Cell Lymphoma Outcomes, Epidemiology and Etiology	Administrative claims data	International Collaboration of Breast Registry Activities (ICOBRA)

such as the transcatheter aortic valve replacement (TAVR) device (designed to repair a narrowed aortic valve via a transcatheter approach) in the national Transcatheter Valve Therapy (TVT) registry; endovascular repair devices for abdominal aortic aneurism via the VISION initiative; and an orthopedic device (OrthoCRN) and other devices within their respective CRNs [59].

In other countries, national registries of procedures involving implanted medical devices have significantly augmented their national surveillance efforts [55–58]. Experience from these international registries used for surveillance and observational research provided valuable insights into the development of CRNs and served as a solid platform for building a novel international surveillance infrastructure for regulatory evidence synthesis (e.g., ICOR and ICVR; see Table 21.3).

Mandated Postmarket Studies

The FDA has a unique statutory authority to mandate postmarket studies, either as a condition of approval or "for cause" later in the postmarket period. For class III devices, the FDA may utilize its premarket approval authority to order studies as a condition of approval. A major regulatory/public health challenge that the FDA is facing is to strike an appropriate balance between pre- and postmarket data collection, such that only data essential to premarket clearance/approval be submitted to ensure reasonable safety and effectiveness, while collecting other data best suited for collection in the postmarket period (such as longitudinal outcomes). Appropriate postmarket questions that can be answered in a mandated postapproval study include long-term safety and effectiveness, real-world experience of the device as it enters broader user populations (clinicians and patients), effectiveness of training programs and learning curve effect, and device performance in certain subgroups of patients not well studied in the premarket clinical trials. Depending on the nature of the postmarket questions, a variety of study designs and approaches can be employed. Designing scientifically sound but practical studies and achieving adequate patient and physician recruitment rates through adequate minimization of loss to follow-up can be particularly challenging for implantable device studies.

In addition, Section 522 of the US Food and Drug Administration Modernization Act (FDAMA) and added regulation (21CFR 822) allows the FDA to require a postmarket surveillance study for class II and III devices (i) that are intended to be implanted in the human body for longer than a year; (ii) that are life sustaining or life supporting (and used outside of the user facility); (iii) whose failure would reasonably be likely to have serious health consequences; or (iv) that are anticipated to have significant use in the pediatric population. Possible study designs vary from detailed review of complaint history and the literature, nonclinical testing, use of registries, observational study designs, and randomized clinical trials [86]. These studies may be ordered at the time of device clearance or approval, or "for cause" later in the postmarket period.

In a series of FDA public workshops held in 2012, the FDA unveiled its National Medical Device Postmarket Surveillance Plan. In 2014, the National Medical Device Postmarket Surveillance System Planning Board and the National Medical Device Registries Task Force were established to guide implementation of this national plan, ultimately leading to the establishment of NEST. The importance of these efforts is that other sources of data, such as national registries, electronic health records, and administrative claims data, are increasingly being used in the conduct of mandated studies in lieu of traditional, one-off studies that require *de novo* data collection.

Administrative Claims Data

Claims and administrative databases have been used to evaluate devices and procedures in multiple clinical areas (see also Chapter 12). The scope

of research ranges from comparative effectiveness to longitudinal assessment of safety events. Results of these studies provide critical information and evidence for physicians and patients to facilitate informed clinical decision-making.

Because administrative data often cover a large and representative population, studies based on these data can provide robust estimates and overviews of population trends in device and procedure use, although not specific to the brand level. Studies based on national and regional databases investigating population use of a novel device or surgical technique reflect the real-world adoption of devices after commercialization. These trends in device use, often assessed with trends in population disease status [87] or healthcare costs [88], can provide a comprehensive evaluation of the dissemination of new technologies and associated practice or population disease change. Furthermore, the representativeness and generalizability of administrative databases make them good candidates for assessing policy-related changes, such as practice change following regulatory activities [89].

Administrative data have also been utilized in the comparative effectiveness assessment of devices and procedures in surgical care. A variety of new devices and techniques have been increasingly adopted in surgical cancer care, including minimally invasive and robot-assisted surgeries and other treatment modalities. National and regional administrative databases and linked databases of administrative data and cancer registries are useful data sources for this type of research. The scope of these studies range from short-term effectiveness, including postoperative complications, length of stay, and readmissions, to long-term safety and effectiveness, such as healthcare costs and patient survival [90]. Apart from the comparison between surgical approaches, the comparative effectiveness of cancer surgery and other treatment modalities such as radiation therapy [91] and stent implantation [92] can also be assessed using these databases.

The usefulness of comparative effectiveness research utilizing administrative data is not limited to cancer surgical care. This type of study is useful in assessing real-world outcomes for certain devices. Administrative data can provide a unique opportunity to evaluate devices such as surgical mesh for pelvic organ prolapse [93] and implants used in hysteroscopic sterilization [94].

Some national and regional databases not only provide representative cross-section estimates, they also have the ability to follow patients up longitudinally and across facilities. For these reasons, administrative data can be used to study safety patterns of permanently implanted devices over the long term [95]. Moreover, some administrative bases record the providers that performed the procedures, making it possible to evaluate provider-level (hospital-level or surgeon-level) effects on patient and device outcomes [96].

The use of administrative databases for epidemiologic research has the strengths of studying large numbers of patients with diverse characteristics and a wide variety of clinical practices, as well as the inclusion of longitudinal data from the continuum of clinical care, and a good representation of vulnerable populations. All these features lead to increased external validity (generalizability) [97] (see Part IIIb). The large number of diverse patients presents opportunities to study treatment effect heterogeneity and to advance methods such as high-dimensional propensity scores and instrumental variables (see also Chapter 43). With regard to medical devices, the limitations of administrative databases include the lack of UDIs, potential inaccuracy of coding of diagnosis, and missing data on laterality and type of revision procedure performed. The lack of clinical information in the administrative billing data can be supplemented by linking the billing data to data from registries or other clinically rich data from other data sources. The FDA has used these data to estimate patient characteristics and in-hospital

mortality rates associated with aortic valve replacements and pacemaker implantation [98,99]. Other researchers have used the data to perform studies on artificial hips [100–104].

Methods for Implantable Device Outcome Evaluation

A methodologic framework for implantable device epidemiology and surveillance involves understanding factors affecting the decision to implant the device, identifying the comparison group(s), and estimating the safety and effectiveness of the device compared to the alternative strategies. In the context of the multiple clinical issues and methodologic challenges noted previously, we believe that a key issue in addressing these goals relates to the multiple sources of variability that exist with implantable devices. These sources relate to systematic and random variations due to the patient, to the surgeon/operator and the center, and to the device itself.

Sources of Variation

Patient Variation (X)

Measurable patient characteristics may predict what type of device is received as well as clinical and device outcomes. For instance, in the case of total hip replacements [104], it has been reported that advanced age, co-morbidities such as heart failure and diabetes, and nonelective admissions were associated with inferior patient outcomes. However, advanced age is also associated with increased use of metal-on-polyethylene hip systems compared to hip systems constructed from other bearing surfaces [103]. The primary reason for the device implant also drives the clinical endpoints. For example, a left-ventricular assist device could be implanted as a "bridge-to-transplant" or as "destination" therapy. In the former, the implant is meant as a relatively short-term solution, whereas for the latter, the implant is meant as a permanent therapy.

Surgeon/Operator and Center Variation (Z)

Surgeon and surgical center skills may have a large impact on the type of hip replacement surgery and clinical outcomes. Several features of the surgical procedure in which the device is implanted vary. For example, orthopedic surgeons may opt to use tissue-sparing surgery when implanting a total hip replacement system. This technique, which differs from the standard lateral direct Hardinge approach, involves smaller incisions and less tissue disruption, which are associated with less pain, reduced blood loss, and shorter hospital stays. However, complications can increase if the surgeon is still early in his/her learning curve [105]. The surgical volume of the surgeon and of the center relate to procedural success. Other features of the surgery can affect the clinical success of the procedure. For example, computer-assisted navigation can increase the accuracy of the positioning of the device.

Device Variation (D)

Several measurable characteristics of devices have been shown to be predictive of device use and outcome. Returning to hip replacement systems, the type of bearing surface is related to revision rates. In particular, hard femoral head and hard cups, such as metal-on-metal or ceramic-on-ceramic, result in lower wear rates. Additionally, large-diameter femoral head size may result in lower dislocation rates. The process of implantation fixation to the bone also results in variations in clinical outcomes. Hip systems can be implanted with bone cement that helps position the implant within the bone, or the systems may have a porous surface that permits bone to grow into its surface.

Understanding the Treatment Assignment Mechanism

A key principle of utilizing observational data for safety surveillance is to view observational studies as approximations to randomized studies.

As such, the first step involves determining how devices are "assigned" to patients. In an RCT the investigator has control over the assignment, whereas in the surveillance setting this mechanism must be estimated. A statistical model uses baseline patient and surgeon characteristics to predict who gets which device. The type of statistical model and type of predictors will depend on the number of alternative competing therapies. For example, if the competing therapy is not a device, patient and surgeon characteristics must be commonly identified predictors. In comparing multiple devices with multiple nondevices, treatment and surgeon characteristics within each type of treatment strategy can vary.

In the case of two different devices, a natural choice to model device selection is a logistic regression that accounts for surgeon-specific random effects. With more than two devices, extensions using multinomial or nested logit models can be estimated [106]. In the latter case, device characteristics can be included to differentiate among devices within a similar class. For example, the treatment options may include the use of a number of different hard-on-hard total hip replacement systems compared to other types of hip replacement systems, such as metal-on-polyethylene. Center-specific effects can be included through the incorporation of an additional variance component. An understanding of the factors associated with device use will identify comparable patients in terms of measurable characteristics for estimating safety and effectiveness for particular cohorts of patients.

Estimating Comparative Effectiveness and Safety

Estimation of the treatment assignment algorithm provides a numeric score for each patient (i.e., the propensity score, or probability of getting the particular device versus the alternative device), and this may be used to validate assumptions required to estimate causal effects. These assumptions relate to unmeasured confounding, positivity of device assignments, and additive device effects. The positivity assumption [107] asserts that an individual subject is eligible to receive all the devices under study. For example, if a device comes in different sizes, some subjects may never receive a particular type of device, as the device is simply too big. In this situation, the positivity condition is violated and the researcher would need to eliminate the subject from the comparison as a causal effect, as the subject is not defined or stratified by device size. The additive treatment effect assumes that the outcome for a subject implanted with device A differs by a constant amount from the outcome had the subject been implanted with device B, and effectively implies no treatment modification by patient subgroups. The definition of a favorable device effect if the device effect is *not* additive requires stating the effect in different patient subgroups.

Semi-parametric estimators, such as matching estimators, weighted estimators, or double-robust estimators that augment weighted estimators with regression estimators [108], can then be used to make inferences. Importantly, with high-dimensional data, machine learning algorithms modified to adhere to assumptions for causal inference can often provide more robust conclusions than standard propensity-score approaches. For example, targeted maximum likelihood estimation is an approach that focuses on estimating a specific parameter rather than the entire likelihood and boasts many desirable statistical properties [109].

Simultaneously Combining All Available Evidence

Assume that rather than a single study, many studies involving a device are available. These studies may reflect different populations, such as registries from different countries or from similar patient populations, or different comparison groups. Novel methods involve assembling all

the available evidence in order to reduce uncertainty about the performance of any particular device. To do this requires assuming that particular relationships among the devices exist, although uncertainty about the relevance of these relationships remains. These uncertainties may relate to how the performance characteristics of different devices relate to patient outcomes; how devices that have been compared in other studies on *similar* outcomes but not to each other are related; and how devices that have been compared in other studies on *different* outcomes but not to each other are related. The appendix to this chapter provides details regarding how to combine the data.

The Future

Device Epidemiology and Digital Health

The broad scope of digital health includes categories such as mobile health (mHealth), health information technology (IT), wearable devices, telehealth and telemedicine, and personalized medicine. These technologies open new opportunities for patients and consumers to better manage and track their health and wellness through greater access to information. The interface of these devices and healthcare will continue to offer new opportunities for medical device epidemiology to lead evidence generation, synthesis, and appraisal.

Translational Epidemiology

Epidemiologic data have an enormous potential to help guide basic science investigations (e.g., guiding the development of biomarkers for detection of patient risks for development of adverse responses to implantable devices). In addition, when combined with preclinical and other data sources (e.g., genetic, explant retrieval), epidemiologic findings could significantly advance evidence generation. In addition, epidemiology could leverage preexisting implant-related data from observational data sources, in individuals with and without implant-related adverse outcomes, to improve our understanding of implant safety and effectiveness. These types of interdisciplinary application of epidemiology could lead to more effective identification of candidate biomarkers predictive of certain implant-related responses (both local and systemic) in different patient subgroups. For instance, *in silico* approaches could combine epidemiologic and other data sources to help guide the development of biomarkers.

Device Epidemiology and Evidence-Informed Practice and Policy

Medical device epidemiology will continue to draw from advances in electronic health records, electronic data capture, standard taxonomy, global patient identifiers, integrated security, and privacy services. Thus, contemporary device epidemiology will be able to mobilize the advances of translational health research sciences through new methods that combine basic science and clinical data, leading to evidence about the choices of best available treatment targeted to specific populations. In doing so, the epidemiology will have to balance the strengths and limitations of systematic review, quantitative, interpretive, narrative, sequenced, and other synthesis approaches within the context of specific public health policy and healthcare settings,

Device Epidemiology and Regulatory Science

MDEpiNet is a global public–private partnership that seeks to advance the collection and use of real-world data to improve patient outcomes. It brings together stakeholders from across the health ecosystem to develop and improve real-world data infrastructure, and carry out studies to better understand how devices perform in

the real world. This international effort is uniquely focused on medical devices and it comes at a particularly opportune time, when many recent developments, ranging from the creation and expansion of device registries to significant strides toward the universal adoption of electronic health records, provide new and promising opportunities for the epidemiologic study of medical devices. The intent is to have a comprehensive, up-to-date risk–benefit profile of specific medical devices at any point in their life cycle, so that optimally informed decisions can be made and provide more useful information to practitioners, patients, and industry. Evolution of public–private partnerships including NEST in the US will drive collaborative knowledge-sharing between members of the ecosystem. Important challenges for MDEpiNet will be to develop and test novel methods for the synthesis and systematic evaluation of all available evidence relevant to a device's risk–benefit profile, including premarket bench, animal, and clinical studies; postmarket surveillance studies; and adverse event reports to advance regulatory needs.

Device Epidemiology and International Infrastructure

The accelerating pace of emerging medical technologies worldwide will continue, and the information science applications are expected to further shape IT-based healthcare, dealing with new demands for storage, transmission, management, and analysis of patient data. The future global impact of epidemiology on our understanding of implantable devices will depend on technological and policy solutions for international collaboration to achieve consistency between global data sources, regulations, and methodologic approaches for various medical device implant applications.

Collaborative research efforts can particularly help to fill a major gap via international consortia. Examples of such collaborative effort are ICOR [110], ICVR [111], and the International Coalition of Breast Registries Associations (I-COBRA) [112]. The development of an international infrastructure creates opportunities for novel method developments for epidemiologic studies. The methods for harmonization, sharing, and combining data are not well developed and require innovative approaches. Such international collaborations, coupled with increasing regulatory convergence driven by IMDRF, present an unprecedented opportunity for influencing clinical and policy decision-making, with enormous public health implications.

Appendix: Simultaneously Combining All Available Evidence

Assume that Y_{sjkm} denotes the m^{th} outcome associated with the k^{th} device for the j^{th} group of patients within the s^{th} study. A model that reflects heterogeneity among the outcomes assumes that the expected or average outcome, denoted $E(Y_{sjkm})$, can be modeled linearly using a link function g(.), generically as

$$g\left(E\left(Y_{sjkm}\right)\right) = \alpha_m + \beta_k + \gamma_{mk} + a_s + b_{j(s)} + c_{k(s)} + d_{m(s)}$$

where α_m = average for m^{th} outcome; β_k = effect of k^{th} device in the average study and for the average outcome; γ_{mk} = deviation from the average of device *k* on outcome *m*; a_s = main effect of s^{th} study; $b_{j(s)}$ = study-specific effect of j^{th} group within s^{th} study; $c_{k(s)}$ = study-specific effect of treatment *k* within s^{th} study; and $d_{m(s)}$ = study-specific effect of outcome *m* within the s^{th} study.

The observed outcomes are summaries, e.g., the average failure rate, rather than at the patient level, although it is a simple modification to include patient-level data. The model assumes that the observed outcomes are connected, in that any observed outcome defined

by (s,j,k,m) is related to or "reached" from any other outcome defined by (s*,j*,k*,m*). This assumption permits borrowing of information from like-devices studies to better estimate the performance of particular devices. Heterogeneity among outcomes is permitted by assuming that γ_{mk}, a_s, $b_{j(s)}$, $c_{k(s)}$, and $d_{m(s)}$ are random effects. Fixed characteristics of the device, D, and of the patient groups, X, can be easily included in the model. Expected differences in outcomes for one device compared to another device averaged over all patient groups can be obtained as functions of the parameters in the model.

References

1 Kalorama Information. *Global Market for Medical Devices*, 8th edn. https://kaloramainformation.com/product/the-global-market-for-medical-devices-8th-edition-40-specific-device-markets-and-23-country-markets-forecasted-to-2022 (accessed May 2019).

2 US Food and Drug Administration. FDA permits marketing of artificial intelligence-based device. April 11, 2018. https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm (accessed May 2018).

3 Federal Food, Drug and Cosmetic Act. Section 201, Definitions (h).

4 World Health Organization. Medical device regulations: global overview and guiding principles. 2003. http://www.who.int/medical_devices/publications/en/MD_Regulations.pdf (accessed May 2018).

5 Council of the European Communities. Directive 93/42/EEC of 14 June 1993 concerning medical devices. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31993L0042:en:HTML (accessed May 2018).

6 Commonwealth of Australia. Therapeutic Goods Act. http://www.comlaw.gov.au/ComLaw (accessed May 2018).

7 Commonwealth of Australia. Therapeutic Goods (Medical Devices) Regulations. http://www.comlaw.gov.au/ComLaw (accessed December 2010).

8 Health Canada. Therapeutic Products Directorate. http://www.hc-sc.gc.ca/contact/dhp-mps/hpfb-dgpsa/tpd-dpt-eng.php (accessed May 2018).

9 International Medical Device Regulators Forum. http://www.imdrf.org/index.asp (accessed May 2018).

10 IMDRF Patient Registries Working Group. Principles of international system of registries linked to other data sources and tools. 2016. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-160930-principles-system-registries.pdf (accessed May 2018).

11 IMDRF Patient Registries Working Group. Tools for assessing the usability of registries in support of regulatory decision-making. 2017. http://www.imdrf.org/consultations/cons-registries-n46-pd1-170817.asp (accessed May 2019).

12 IMDRF Patient Registries Working Group. Methodological principles in the use of international medical device registry data. 2016. http://www.imdrf.org/consultations/cons-methodological-principles-imdrd.asp (accessed May 2018).

13 US Food and Drugs Administration. Use of real-world evidence to support regulatory decision-making for medical devices. 2017. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM513027.pdf (accessed May 2018).

14 US Food and Drugs Administration. Balancing premarket and postmarket data collection for devices subject to premarket approval. 2015. https://www.fda.gov/regulatory-information/

search-fda-guidance-documents/balancing-premarket-and-postmarket-data-collection-devices-subject-premarket-approval (accessed May 2019).

15 US Food and Drugs Administration. Factors to consider when making benefit–risk determinations in medical device premarket approval and de novo classifications. 2016. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/factors-consider-when-making-benefit-risk-determinations-medical-device-premarket-approval-and-de (accessed May 2019).

16 US Food and Drugs Administration. Patient preference information – voluntary submission, review in premarket approval applications, humanitarian device exemption applications, and de novo requests, and inclusion in decision summaries and device labeling. 2016. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-preference-information-voluntary-submission-review-premarket-approval-applications (accessed May 2019).

17 European Commission. MEDDEV 2.7.1. rev. 4. Clinical evaluation: a guide for manufacturers and notified bodies. http://ec.europa.eu/consumers/sectors/medicaldevices/files/meddev/2_4_1_rev_9_classification_en.pdf (accessed May 2018).

18 Australian Government. Registration or listing of therapeutic goods. https://ablis.business.gov.au/service/ag/registration-or-listing-of-therapeutic-goods/3 (accessed May 2019).

19 Muni NI, Gross TP, Boam AB, Wang S, Zuckerman BD. Challenges in regulating breakthrough medical devices. *Food Drug Law J* 2005; **60**(2): 137–42.

20 Roback K, Gaddlin PO, Nelson N, Persson J. Adoption of medical devices: perspectives of professionals in Swedish neonatal intensive care. *Technol Health Care* 2007; **15**(3): 157–79.

21 Hammill S, Phurrough S, Brindis R. The National ICD Registry: now and into the future. *Heart Rhythm* 2006; **3**(4): 470–3.

22 Ramsay CR, Wallace SA, Garthwaite PH, Monk AF, Russell IT, Grant AM. Assessing the learning curve effect in health technologies: lessons from the nonclinical literature. *Int J Technol Assess Health Care* 2002; **18**(1): 1–10.

23 Hannan EL, Racz M, Ryan TJ, *et al*. Coronary angioplasty volume–outcome relationships for hospitals and cardiologists. *JAMA* 1997; **277**(11): 892–8.

24 Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med* 2003; **349**(22): 2117–27.

25 Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002; **137**(6): 511–20.

26 Holt PJ, Poloniecki JD, Loftus IM, Thompson MM. The relationship between hospital case volume and outcome from carotid endartectomy in England from 2000 to 2005. *Eur J Vasc Endovasc Surg* 2007; **34**(6): 646–54.

27 Nordin P, van der Linden W. Volume of procedures and risk of recurrence after repair of groin hernia: national register study. *BMJ* 2008; **336**(7650): 934–7.

28 Resnic FS, Wang TY, Arora N, *et al*. Quantifying the learning curve in the use of a novel vascular closure device: an analysis of the NCDR (National Cardiovascular Data Registry) CathPCI registry. *JACC Cardiovasc Interv* 2012; **5**(1): 82–9. doi: 10.1016/j.jcin.2011.09.017

29 Rosenfield KM. Clinical competence statement on carotid stenting: training and credentialing for carotid stenting – multispecialty consensus recommendations. *J Vasc Surg* 2005; **41**(1): 160–8.

30 Tavris D, Shoaibi A, Chen AY, Uchida T, Roe MT, Chen J. Gender differences in the treatment of non-ST-segment elevation myocardial infarction. *Clin Cardiol* 2010; **33**(2): 99–103.

31 Kaczmarek RG, Liu CH, Gross TP. Medical device surveillance: gender differences in

pulmonary artery rupture after pulmonary artery catheterization. *J Womens Health (Larchmt)* 2003; **12**(9): 931–5.

32 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342**(25): 1887–92.

33 Dreyer NA, Garner S. Patient registries, predictive models and optimal care. *JAMA* 2009; **302**(24): 2662.

34 American Recovery and Reinvestment Act (ARRA), Public Law (111-5). 2009.

35 Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: IOM, 2009.

36 Chalkidou K, Tunis S, Lopert R, *et al*. Comparative effectiveness research and evidence-based health policy: experience from four countries. *Milbank Q* 2009; **87**(2): 339–67.

37 Normand SL, Marinac-Dabic D, Sedrakyan A, Kaczmarek R. Rethinking analytical strategies for surveillance of medical devices: the case of hip arthroplasty. *Med Care* 2010; **48**(6 Suppl): S58–S67.

38 Sedrakyan A, Marinac-Dabic D, Normand SL, Mushlin A, Gross T. A framework for evidence evaluation and methodological issues in implantable device studies. *Med Care* 2010. **48**(6 Suppl): S121–S128.

39 Marinac-Dabic D, Bonhomme M, Loyo-Berrios N, *et al*. Medical devices post-approval studies program: vision, strategies, challenges and opportunities. *Food Drug Law J* 2007; **62**(3): 597–604.

40 Swedish National Joint Registry. Swedish Knee Arthroplasty Register. http://www.myknee.se/en (accessed May 2018).

41 National Joint Registry. Welcome from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. http://www.njrcentre.org.uk/njrcentre/default.aspx (accessed May 2018).

42 Australian Orthopaedic Association National Joint Replacement Registry. Welcome to the AOANJRR. https://aoanjrr.sahmri.com/en (accessed May 2018).

43 Canadian Institute for Health Information. Canadian Joint Replacement Registry (CJRR). http://www.cihi.ca/CIHI-ext-portal/internet/EN/TabbedContent/types+of+care/specialized+services/joint+replacements/cihi021359 (accessed May 2018).

44 Medical Device Epidemiology Network. Building global real world evidence collaboratives for health technologies. http://mdepinet.org (accessed May 2018).

45 US Food and Drugs Administration. Unique Device Identification – UDI. 2018. https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/UniqueDeviceIdentification/default.htm (accessed May 2018).

46 Bright R. Special applications of pharmacoepidemiology. In Strom B, Kimmel S, eds., *Textbook of Pharmacoepidemiology*. Chichester: John Wiley & Sons, 2007, pp. 339–406.

47 Medical Device Innovation Consortium. Advancing the health and safety of patients through innovation. http://mdic.org (accessed May 2018).

48 Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000; **342**(25): 1907–9.

49 Kunz L, Rose S, Spiegelman D., Normand S-L. An overview of statistical approaches for comparative effectiveness research. In Gatsonis C, Morton S (eds), *Methods in Comparative Effectiveness Research*. New York: CRC Press, 2017, pp. 1–35.

50 Normand SL. Some old and some new statistical tools for outcomes research. *Circulation* 2008; **118**(8): 872–84.

51 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**(8 Pt 2): 757–63.

52 D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; **17**(19): 2265–81.

53 Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assn* 1996; **91**: 444–72.

54 Heckman JJ. Comment on "Identification of causal effects using instrumental variables." *J Am Stat Assn* 1996; **91**: 459–62.

55 Basu A. Estimating decision-relevant comparative effects using instrumental variables. *Stat Biosci* 2011; **3**(1): 6–27.

56 Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc B* 1996; **58**(1): 267–88.

57 Ravikumar P, Lafferty J, Liu H, Wasserman L. Sparse additive models. *J Royal Stat Soc B* 2009; **71**: 1009–30.

58 Gelman A, Jakulin A, Grazia P, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2008; **2**(4): 536–49.

59 Mccaffrey D, Griffin BA, Almirall D, Slaughter M, Ramchand R, Burgette LD. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat Med* 2013; **32**: 3388–414.

60 van der Laan M, Rose S. *Targeted Learning: Prediction and Causal Inference for Observational and Experimental Data*. New York: Springer, 2011.

61 US Food and Drug Administration. Medwatch Form 3500A. 2015. https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Forms/UCM048334.pdf (accessed May 2018).

62 US Food and Drug Administration. MAUDE – Manufacturer and User Facility Device Experience. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm (accessed May 2018).

63 Medical device problem reporting for the betterment of healthcare. *Health Devices* 1998; **27**(8): 277–92.

64 DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System (with discussion). *Am Stat* 1999; **53**(3): 177–90.

65 DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM, pp. 67–76.

66 US Food and Drug Administration. MDR adverse event codes. 2018. https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/ReportingAdverseEvents/MDRAdverseEventCodes/default.htm (accessed May 2018).

67 US Food and Drug Administration. MedSun: Medical Product Safety Network. 2018. https://www.fda.gov/medical-devices/medical-device-safety/medsun-medical-product-safety-network (accessed May 2019).

68 Bright RA, Flack MN, Gardner SN. The Medical Product Surveillance Network (MedSun). In Brown SL, Bright RA, Tavris DR, eds., *Medical Device Epidemiology and Surveillance*. Chichester: John Wiley & Sons, 2007, pp. 63–77.

69 FDA Public Health Notification: Serious complications associated with transvaginal placement of surgical mesh in repair of pelvic organ prolapse and stress urinary incontinence. 2011. http://wayback.archive-it.org/7993/20170722150848/ https://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm262435.htm (accessed May 2018).

70 FDA Public Health Notification: Life-threatening complications associated with recombinant human bone morphogenetic protein in cervical spine fusion. 2008. https://wayback.archive-it.org/7993/20170112170147/http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm079019.htm (accessed May 2018).

71 FDA Public Health Notification: MRI-caused injuries in patients with implanted neurological stimulators. 2005. https://wayback.archive-it.

org/7993/20170111190630/http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/PublicHealthNotifications/ucm062125.htm (accessed May 2018).

72 Reefhuis J, Honein MA, Whitney CG, *et al*. Risk of bacterial meningitis in children with cochlear implants. *N Engl J Med* 2003; **349**(5): 435–45.

73 Tavris DR, Dey S, Albrecht-Gallauresi B, *et al*. Risk of local adverse events following cardiac catheterization by hemostasis device use – phase II. *J Invasive Cardiol* 2005; **17**(12): 644–50.

74 Tavris DR, Gallauresi BA, Lin B, *et al*. Risk of local adverse events following cardiac catheterization by hemostasis device use and gender. *J Invasive Cardiol* 2004; **16**(9): 459–64.

75 Sedrakyan A, Paxton E, Graves S, Love R, Marinac-Dabic D. National and international postmarket research and surveillance implementation: achievements of the International Consortium of Orthopaedic Registries initiative. *J Bone Joint Surg Am* 2014; **96**(Suppl 1): 1–6.

76 Macpherson GJ, Brenkel IJ, Smith R, Howie CR. Outlier analysis in orthopaedics: use of CUSUM: the Scottish Arthroplasty Project: shouldering the burden of improvement. *J Bone Joint Surg Am* 2011; **93**(Suppl 3): 81–8.

77 de Steiger FR, Hang JR, Miller LN, Graves SE, Davidson DC; FRACS. Five-year results of the ASR Xl acetabular system and the ASR hip resurfacing system: an analysis from the Australian Orthopaedic Association National Joint Replacement Registry. *J Bone Joint Surg Am* 2011; **93**(24): 2287–93.

78 Shewhart W. The application of statistics as an aid in maintaining quality of a manufactured product. *J Am Stat Assn* 2012; **1925**(20): 546–8.

79 Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005; **24**: 1185–202.

80 Matheny ME, Morrow DA, Ohno-Machado L, Cannon CP, Sabatine MS, Resnic FS. Validation of an automated safety surveillance system with prospective, randomized trial data. *Med Decis Making* 2009; **29**(2): 247–56.

81 Resnic FS, Gross TP, Marinac-Dabic D, *et al*. Automated surveillance to detect postprocedure safety signals of approved cardiovascular devices. *JAMA* **304**(18): 2019–27.

82 Dusetzina SB, Tyree S, Meyer AM, *et al*. Linking data for health services research: a framework and instructional guide. Report No. 14-EHC033-EF. Rockville, MD: Agency for Healthcare Research and Quality, 2014. https://www.ncbi.nlm.nih.gov/pubmed/25392892 (accessed May 2018).

83 Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995; **14**(5–7): 491–8.

84 Sedrakyan A, Campbell B, Graves S, Cronenwett JL. Surgical registries for advancing quality and device surveillance. *Lancet* 2016; **388**(10052): 1358–60. doi:10.1016/S0140-6736(16)31402-7

85 US Food and Drug Administration. Recommendations for a national medical device evaluation system: strategically coordinated registry networks to bridge clinical care and research. 2015. http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM459368.pdf (accessed May 2018).

86 CFR – Code of Federal Regulations Title 21, Pt. 822, 2002 ed. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=822 (accessed May 2018).

87 Jalbert JJ, Isaacs AJ, Kamel H, Sedrakyan A. Clipping and coiling of unruptured aneurysms among Medicare beneficiaries, 2000 to 2010. *Stroke* 2015; **46**(9): 2452–7.

88 Salzler GG, Meltzer AJ, Mao J, *et al*. Characterizing the evolution of perioperative outcomes and costs of endovascular abdominal aortic aneurysm repair. *J Vasc Surg* 2015; **62**(5): 1134–9.

89 Sedrakyan A, Chughtai B, Mao J. Regulatory warnings and use of surgical mesh in pelvic

organ prolapse. *JAMA Intern Med* 2016; **176**(2): 275–7.

90 Hu JC, Chughtai B, O'Malley P, *et al*. Perioperative outcomes, health care costs, and survival after robotic-assisted versus open radical cystectomy: a national comparative effectiveness study. *Eur Urol* 2016; **70**(1): 195–202.

91 Paul S, Lee PC, Mao J, Isaacs AJ, Sedrakyan A. Long term survival with stereotactic ablative radiotherapy (SABR) versus thoracoscopic sublobar lung resection in elderly people: national population based study with propensity matched comparative analysis. *BMJ* 2016; **354**: i3570.

92 Abelson JS, Yeo HL, Mao J, Milsom JW, Sedrakyan A. Long-term postprocedural outcomes of palliative emergency stenting vs stoma in malignant large-bowel obstruction. *JAMA Surg* 2017; **152**(5): 429–35.

93 Chughtai B, Mao J, Buck J, Kaplan S, Sedrakyan A. Use and risks of surgical mesh for pelvic organ prolapse surgery in women in New York state: population based cohort study. *BMJ* 2015; **350**: h2685.

94 Mao J, Pfeifer S, Zheng XE, Schlegel P, Sedrakyan A. Population-based estimates of the prevalence of uterine sarcoma among patients with leiomyomata undergoing surgical treatment. *JAMA Surg* 2015; **150**(4): 368–70.

95 Chughtai B, Thomas D, Sun T, Sedrakyan A. Failures of sacral neuromodulation for incontinence. *JAMA Surg* 2018; **Feb 14**: e176093.

96 Meltzer AJ, Agrusa C, Connolly PH, Schneider DB, Sedrakyan A. Impact of provider characteristics on outcomes of carotid endarterectomy for asymptomatic carotid stenosis in New York State. *Ann Vasc Surg* 2017; **45**: 56–61.

97 Malenka DJ, Kaplan AV, Sharp SM, Wennberg JE. Postmarketing surveillance of medical devices using Medicare claims. *Health Aff (Millwood)* 2005; **24**(4): 928–37.

98 Astor BC, Kaczmarek RG, Hefflin B, Daley WR. Mortality after aortic valve replacement: results from a nationally representative database. *Ann Thorac Surg* 2000; **70**(6): 1939–45.

99 Daley WR, Kaczmarek RG. The epidemiology of cardiac pacemakers in the older US population. *J Am Geriatr Soc* 1998; **46**(8): 1016–19.

100 Greenland S, Finkle WD. A case-control study of prosthetic implants and selected chronic diseases in Medicare claims data. *Ann Epidemiol* 1998; **8**(5): 319–26.

101 Greenland S, Finkle WD. A retrospective cohort study of implanted medical devices and selected chronic diseases in Medicare claims data. *Ann Epidemiol* 2000; **10**(4): 205–13.

102 Fisher ES, Baron JA, Malenka DJ, Barrett J, Bubolz TA. Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. *Am J Public Health* 1990; **80**(12): 1487–90.

103 Bozic KJ, Ong K, Lau E, *et al*. Risk of complication and revision total hip arthroplasty among Medicare patients with different bearing surfaces. *Clin Orthop Relat Res* 2010; **468**(9): 2357–62.

104 Zhan C, Kaczmarek R, Loyo-Berrios N, Sangl J, Bright RA. Incidence and short-term outcomes of primary and revision hip replacement in the United States. *J Bone Joint Surg Am* 2007; **89**(3): 526–33.

105 D'Arrigo C, Speranza A, Monaco E, Carcangiu A, Ferretti A. Learning curve in tissue sparing total hip replacement: comparison between different approaches. *J Orthop Traumatol* 2009; **10**(1): 47–54.

106 Tchernis R, Horvitz-Lennon M, Normand SL. On the use of discrete choice models for causal inference. *Stat Med* 2005; **24**(14): 2197–212.

107 Rosenabaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.

108 Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a

comparative study. *Stat Med* 2004; **23**(19): 2937–60.

109 Van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Berlin: Springer, 2011.

110 International Consortium of Orthopedic Registries (ICOR). ICOR-USA Collaboration. http://www.icor-initiative.org/icor-usa.aspx (accessed May 2018).

111 International Consortium of Vascular Registries (ICVR). http://icvr-initiative.org (accessed May 2018).

112 Australian Society of Plastic Surgeons. International Collaboration of Breast Registry Activities (ICOBRA). https://plasticsurgery.org.au/protecting-patient-safety/icobra (accessed May 2018).

22

Research on the Effects of Medications in Pregnancy and in Children

Daniel B. Horton[1], Sonia Hernandez-Diaz[2], Tamar Lasky[3], and Krista F. Huybrechts[4]

[1] *Department of Pediatrics, Rutgers Robert Wood Johnson Medical School, Rutgers Center for Pharmacoepidemiology and Treatment Science, New Brunswick, NJ, USA*
[2] *Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA*
[3] *Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA*
[4] *Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

For ethical and other reasons, pregnant women and children have traditionally been excluded from premarketing clinical trials, under the assumption that findings of efficacy and safety can be extrapolated from nonpregnant adult populations. These well-intentioned protections from the risks and burdens of trials have paradoxically left pregnant women and children more susceptible to the potential risks of drugs used in everyday practice, frequently off-label and without high-quality evidence of efficacy, effectiveness, or safety. The field of pharmacoepidemiology has an important role in supplying critical missing evidence about the beneficial and harmful effects of drugs, both old and new, in pregnant women and children. However, pharmacoepidemiologic research in these populations presents numerous methodologic and practical challenges. Rapid and dramatic changes in growth, development, and physiology during pregnancy, infancy, and childhood can alter drugs' pharmacokinetics, pharmacodynamics, and, thus, efficacy and safety. The specific biology and varying incidence of clinical conditions that affect each stage of pregnancy and childhood also require consideration of separate subgroups. Given the infrequent use of most specific medications in these populations and the rarity of many serious outcomes of interest, problems arise related to sample size. In addition to random error, observational studies in pregnant women and children are subject to systematic errors – some errors that are shared with any pharmacoepidemiologic study and others that are fairly specific to these populations.

Case–control studies can efficiently estimate associations, provided the drugs are relatively commonly used. However, this design faces the challenge of retrospective data collection, selection of valid controls, and samples restricted to motivated volunteers. Exposed pregnancy registries can more efficiently oversample exposed pregnancies for uncommonly used drugs, collect drug utilization before outcomes are known, and estimate absolute risks rather than associations

The authors wish to thank Bruce Carleton, PharmD, Susan dosReis, PhD, Alan C. Kinlaw, PhD, Andrea V. Margulis, MD, ScD, Jonathan Slaughter, MD, MPH, and Rachel E. Sobel, DrPH, for their valuable input and suggestions. This book chapter reflects the views of the authors and should not be construed to represent the FDA's views or policies.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

alone. However, this design still includes populations of volunteers and requires long, costly study periods to recruit and follow sufficient participants. Moreover, case–control designs are restricted to a few outcomes, and pregnancy registries are restricted to a few drugs. Administrative healthcare or clinical databases can provide large study samples to address these difficulties, and pooling efforts across data sources is increasingly common. However, for research in pregnant and pediatric populations, automated databases may lack important dates (for instance, of conception or birth) and other key clinical information (such as children's weight). For other limitations of automated databases, see Chapters 11–14.

Various forms of bias can occur in observational research in pregnant and pediatric populations. Confounding may involve unmeasured factors (e.g., health-seeking behaviors related to prenatal screening or parenting style, environmental exposures, income, genetics) and long latencies between maternal or early life exposures and long-term pediatric outcomes. Moreover, when selecting the study population of pregnant women or children, sources of bias may include depletion of susceptibles (e.g., lethal malformation resulting in undocumented pregnancy losses), left truncation (e.g., enrollment of pregnancies after prenatal screening), right censoring (e.g., exclusion of terminations), and selective inclusion, exclusion, or dropout of vulnerable populations (e.g., children in foster care). Variability in the duration of pregnancy due to health problems may introduce spurious associations due to differential opportunities for drug exposure (e.g., acute drug exposure in the third trimester and a lower risk of preterm delivery) and competing risks (e.g., an increase in spontaneous abortions results in a lower risk of preeclampsia). Importantly, both pregnant and pediatric populations are marked by important genetic and environmental interdependencies: the pregnant woman and the embryo, fetus, and infant; children and their families. These intimate relationships have critical implications for health; information about family members can be instrumental to measuring certain exposures (e.g., prenatal medicines preceding congenital anomalies), controlling for confounding (e.g., family history of psychiatric illness), and identifying effect modifiers. Studies of siblings discordant for prenatal or early life exposures may permit better control for certain unmeasured confounders than traditional cohort studies. Father's medication use can serve as a negative control to differentiate *in utero* drug exposure from familial confounding. Nonetheless, not all data sources have available and valid linkages within families that facilitate such approaches.

This chapter focuses on two populations underrepresented in clinical trials – pregnant women and children – that have specific characteristics necessitating adaptations of study designs and other methodologic considerations. Certain considerations relate to shared interests in safety outcomes in children, as both maternal exposures and childhood exposures can adversely affect pediatric health. We will discuss (i) unique clinical aspects of pregnant and pediatric populations that relate to exposures, outcomes, and confounding; (ii) methodologic issues and considerations for pharmacoepidemiologic research in pregnant women and children; and (iii) potential solutions and settings to study drug utilization and effects in these populations. Each section of the chapter is organized by topic, and each topic contains material specific to pregnant women and/or children, sometimes preceded by information that applies to both populations.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Unique Biology and Epidemiology

Pregnant Women

A woman's body undergoes vast physiologic changes in pregnancy to accommodate and enable

fetal growth, development, and childbirth. These biologic processes lead to rapid changes in baseline risks of maternal conditions and treatment indications that may occur in other stages of life, such as nausea, diabetes mellitus, gallstones, and thromboembolism. However, some outcomes are unique to pregnant women (e.g., ectopic pregnancy, preeclampsia, placental abruption) or their offspring (e.g., neural tube defects, prematurity, cognitive delays). One must consider the fluctuating timelines of risk for outcomes and susceptibility to treatment effects. For example, the risk of pregnancy losses is highest during the first trimester, and the risk of preeclampsia due to endothelial dysfunction increases toward the end of pregnancy. Regarding teratogenicity, while critical portions of organogenesis occur in the first trimester, fetal organs continue to develop later in pregnancy, which explains later susceptibility to certain effects of treatment (e.g., nephrotoxicity from use of angiotensin-converting enzyme inhibitors in the second or third trimester). Furthermore, epigenetic changes and alterations of germline cells can lead to adverse outcomes across generations (e.g., diethylstilbestrol [1,2]), presenting additional challenges in measuring the timing of exposure.

Children

The epidemiology of pediatric diseases varies considerably from that for adults: acute illnesses are considerably more common than chronic diseases; some medical conditions are unique to children, while other conditions seen in adults do not affect children; and some pediatric diseases may have different clinical presentations, treatments, and prognoses from analogous conditions in adults. Considerations of timing of exposure and follow-up are also crucial when studying the treatment of children, who grow and develop dramatically across multiple health axes (e.g., physical, immunologic, nutritional, cognitive, social, behavioral, etc.) and whose own medical needs and risks change throughout childhood. Preterm infants might face any number of life-threatening acute events (e.g., respiratory distress syndrome, necrotizing enterocolitis) and long-term conditions (e.g., retinopathy of prematurity, bronchopulmonary dysplasia) that spare nearly all term infants. These differences underscore the importance of having information such as gestational age at birth, birth weight, and a granular consideration of age when evaluating outcomes in infants. Certain chronic early-onset pediatric diseases, such as type 1 diabetes and some forms of juvenile idiopathic arthritis, differ in incidence, pathophysiology, and clinical course from analogous diseases in adolescents and adults (e.g., type 2 diabetes, rheumatoid arthritis) and thus require independent validation. Other conditions, such as asthma, depression, or infections, may occur in children and adults, but have different treatments or prognoses based on age. For example, compared with adults, children have a substantially greater risk of developing rheumatic heart disease following group A streptococcal pharyngitis [3]. Researchers must consider carefully these age-related clinical differences when designing, conducting, or interpreting pediatric pharmacoepidemiologic research.

Treatment Responses and Patterns

Pregnant Women

The physiologic changes associated with pregnancy can markedly alter pharmacokinetics. Extrapolating conclusions regarding dosing, efficacy, and safety from nonpregnant populations will often be incorrect [4]. For example, the effectiveness of specific drug dosages may be affected by changes in drug clearance, while drug safety for the fetus throughout pregnancy may be affected by changes in the permeability of the blood–placental barrier and may pertain to the ongoing processes of structural development and fetal growth.

Drug utilization patterns (see Chapter 18) and medication adherence (see Chapter 38) also vary

more markedly around pregnancy. By 2008, over 90% of women in the US reported the use of at least one prescription or over-the-counter (OTC) medication during pregnancy, marking a steady increase since the 1970s [5,6]. Some agents are contraindicated (e.g., live vaccines) while others are strongly recommended (e.g., inactivated influenza vaccine). Clinicians frequently change or discontinue chronic medications due to perceived or real risks associated with use during pregnancy (e.g., lithium, valproate) or expected improvement of symptoms during pregnancy (e.g., disease-modifying therapies for multiple sclerosis). Women often decide themselves to discontinue their medications upon learning of their pregnancy because of concerns over fetal toxicity. Treatment decisions during pregnancy are frequently not evidence based and may result in inadequate and ineffective treatments that, nonetheless, expose the fetus to treatment-related risks [6]. Nonadherence or inappropriate discontinuation could also lead to undertreatment of diseases (e.g., asthma, systemic lupus erythematosus) that could adversely affect outcomes of pregnancy, of offspring, and of women after pregnancy [7]. These changing, often unrecorded patterns of maternal drug usage can lead to misclassification of exposure.

Children

As in pregnancy, pharmacokinetics changes rapidly in childhood with the maturation of the various organs needed to absorb, distribute, metabolize, and excrete drugs [8–10]. The pharmacokinetic machinery is particularly underdeveloped in premature infants, for whom there are special considerations in drug dosing [11]. Pharmacodynamics also change over time because of age-related alterations in receptor density and molecular pathways [8,12–14]. Pediatric drugs are generally dosed based on body size, usually weight. Because of developmental changes in metabolism and excretion, in proportion to body size, per-kilogram dosing tends to be lowest in infants (particularly premature infants) and highest in toddlers and younger children; dosing for adolescents tends to be similar to that in adults, but this may not be true for all drugs or all stages of adolescence and requires drug-specific verification [15]. Drug labels for approved pediatric medications specify dosing based on weight and/or age. Guidance for dosing of off-label drugs in children may come from institutional, regional, or national formularies, but this guidance may reflect expert opinion more than sound evidence. The frequency of off-label drug usage in children combined with a lack of pediatric pharmacokinetic/pharmacodynamic data and universal formularies for many pediatric drugs raises questions about both effectiveness and safety.

Rates of medication use among children have changed over time, with rises in the use of drugs for chronic disease such as asthma and attention deficit hyperactivity disorder, and of drug classes such as contraceptives and glucocorticoids [16–19]. About 22–25% of children in the US take a prescribed drug each month [17], while over half of Canadian children receive at least one prescribed medication per year [20]. The prevalence of use of medicines at different ages reflects, in part, the changing prevalence of diseases throughout childhood [16,21]. The ability and willingness of children to take medications also change with age. Young children, for example, may be unable to swallow pills or unwilling to take unpalatable oral medications. Older children and adolescents with more control and independence may also be less adherent to medications or more likely to take medicines that are ineffective or unsafe [22]. Thus, children's adherence, and consequently drug effectiveness, may depend on available formulations, palatability, access (e.g., at school), and the capacity of caregivers to give medication despite children's refusal. Further, parents may inadvertently give incorrect doses of liquid medicines because of confusion about dispensing devices, drug

concentration, or directions [23,24]. These physical measurement issues may not be apparent in either prescribing or dispensing data, but could still compromise the effectiveness or safety of drugs given to children in liquid formulations. Childhood is also a time of experimentation, ranging from the toddler who finds an open medicine bottle to adolescents abusing prescribed or diverted psychotropic or pain medications, leading in some situations to misclassification of exposure or dosage as well as additional concerns for drug safety.

Biologic Plausibility for Teratogenicity: Pregnant Woman

Teratogens are agents that cause birth defects in the embryo (in Greek, *terato* means monster and *genesis* means birth). Whereas one can often predict a drug's adverse effects based on its molecular structure or class and its pharmacologic or toxicologic properties, this is not the case for teratogenic effects. There are a few instances where *in vitro* and animal experiments support the biologic plausibility of drug–defect associations: these include the increased risk of anomalies derived from neural crest cells among infants exposed to retinoids [25] and the decreased risk of neural tube defects among infants exposed to folic acid [26,27]. However, biologic mechanisms remain unknown for many accepted human teratogens, including thalidomide. Thus, we cannot dismiss effects simply because they lack a biologically plausible explanation [28].

Growth and Development: Children

Multiple factors regulate pediatric growth and development, including genetics, hormones, nutrition, and the environment, as well as diseases and their treatments. States and rates of growth and development, thus, provide critical windows into pediatric health. Pediatric health professionals routinely monitor children's growth using national or international growth charts. Absolute measurements can be useful if they deviate markedly from typical values, for example in defining pediatric obesity [29], wasting (low weight), and stunting (low height) [30]. Even when growth measurements are not extreme, changes in growth trajectory can also indicate important changes in health [31,32]. Like growth, there are multiple standards for charting development across multiple axes (e.g., cognitive, motor, social, skeletal, sexual) from early infancy through puberty.

Ethical and Regulatory Context for Research in Pregnant, Lactating, and Pediatric Populations

Regulations are in place throughout the world to guide the approval and surveillance of medications used by pregnant and lactating women and by children. These are based on ethical principles that have been debated and modified over time [33]. In general terms, a historical emphasis on protecting pregnant women and children from research has been counterbalanced by the realization that exclusion of pregnant women and children exposes these populations to treatments with inadequate evidence about efficacy and safety and blocks their access to new medications (e.g., new drugs for resistant tuberculosis undergoing Phase III trials [34]). The ethical concern about protection of pregnant women and children has been based, in part, on their designation as "vulnerable populations." However, the term "vulnerable" is a broad reference to different types of vulnerability and carries varying nuances in different contexts, including different definitions used by regulatory bodies of different countries and organizations [35,36]. As noted by Bracken-Roche, in the case of pregnant women, "vulnerable" historically referred to concerns about drugs that may cross the placenta and adversely affect the fetus; the

term was also an acknowledgment that pregnant women have dual concerns for the welfare of the fetus and themselves [36]. Additional issues may arise in societies where women have limitations on their autonomy. The categorization of pregnant women as vulnerable has been challenged by some bioethicists in ways that might influence our thinking about clinical research in the future [37,38]. In many circumstances, children are considered vulnerable subjects because they are developing the cognitive and emotional abilities to make decisions and to protect themselves, and must rely on someone else (typically a parent or caregiver) who has the capacity to protect their interests [35,36]. Children's vulnerability also relates to increased susceptibility to infectious diseases (including higher risk of mortality at young ages) and to long-term impacts of exposures on health and development [35,36].

Pregnant Women

Regulatory issues around research on medications used by pregnant women are intertwined with general research policies regarding the inclusion of women, pregnant and nonpregnant, in drug trials [39]. Revelations of the teratogenicity of thalidomide led to the passing in the US of the Kefauver–Harris Amendments in 1962 requiring evidence of efficacy for drug approval, but this legislation did not require any evidence for efficacy or safety in pregnancy [40]. Then in 1977, the US Food and Drug Administration (FDA) recommended exclusion of women of childbearing age from participating in Phase I and early Phase II trials. This well-intentioned guideline inadvertently led to the exclusion of the same population from Phase III and IV trials as well. This pattern reversed in 1994 with the establishment of an FDA Office of Women's Health to promote the inclusion of women in clinical trials and the implementation of new policies [41]. However, no mandate emerged to collect clinical trial data specifically in pregnant women.

In 2002, the FDA issued a Guidance to Industry providing recommendations on how to establish pregnancy exposure registries to monitor for outcomes of pregnancies exposed to drug products (see Chapter 16) [42]. In Europe, the 2014 EU Clinical Trial Regulation (No 536/2014) defined the conditions for inclusion of pregnant and lactating women in clinical trials, as well as protective measures to consider for all women of child-bearing age [43,44]. The FDA's Pregnancy and Lactation Labeling Rule (PLLR), effective in 2015, required changes to the content and format of information presented on prescription drug labels, including separate sections for pregnant women and for lactating women [45]. The labeling information is intended to assist clinicians in assessing benefit versus risk and in counseling pregnant women and nursing mothers about medications. The US 21st Century Cures Act (2016) established the Task Force on Research Specific to Pregnant Women and Lactating Women to study the effectiveness and safety of therapies and inform drug regulation in these populations [46]. Additional guidance on research protections of women of child-bearing age has come from the Council for International Organizations of Medical Sciences (CIOMS), a nongovernmental organization. The CIOMS ethical guidelines discuss the importance of informed consent regarding potential risks to a fetus in future pregnancies, as well as access to family planning services (e.g., contraception) and, if needed, obstetric medical care to help women manage unplanned pregnancies [47]. Multiple governments and organizations continue to develop new guidance and resources to address the complexities in this arena.

Children

The CIOMS ethical guidelines note:

> Children and adolescents must be included in health-related research unless a good scientific reason justifies their exclusion. As

> children and adolescents have distinctive physiologies and health needs, they merit special consideration by researchers and research ethics committees. However, their distinctive physiologies and emotional development may also place children and adolescents at increased risk of being harmed in the conduct of research. Moreover, without appropriate support, they may not be able to protect their own interests due to their evolving capacity to give informed consent. Specific protections to safeguard children's rights and welfare in the research are therefore necessary. [47]

The International Conference on Harmonization of Technical Requirements of Pharmaceuticals for Human Use (ICH) has worked to harmonize international policies on pediatric clinical trials [48]. Both Europe and the US have regulations in place requiring clinical data from pediatric populations; these regulations balance the ethical imperative to protect children alongside the ethical need to provide children with treatments that have been adequately tested. In the US, the 2003 Pediatric Research Equity Act authorized the FDA to require pediatric studies of drugs or biologics if the product was likely to be used in a substantial number of pediatric patients, or if it would provide meaningful benefits for children over existing treatments [49]. In Europe, the 2007 Pediatric Regulation dramatically changed the regulatory environment for pediatric medicines [50]. This regulation established the Pediatric Committee (PDCO), which determines the studies that companies must conduct in children as part of pediatric investigation plans (PIPs). The PIP covers the studies needed to support a pediatric indication with an age-appropriate formulation. In addition to regulations requiring studies in children, legislation such as the Food and Drug Administration Modernization Act (1997) and the Best Pharmaceuticals for Children Act (2002) has provided financial incentives (in the form of patent extensions) to pharmaceutical companies that study medications in children. Such legislation has led to increases in pediatric drug research and approvals over time, even though many drugs used in children (including orphan drugs for rare diseases) continue to lack pediatric labeling [51–53].

Evidence to Inform Clinical Practice and the Role of Pharmacoepidemiology

There is usually very little premarketing information on the safety and efficacy of drugs in pregnant or pediatric populations. Given that a drug's safety can rarely be predicted based on its structure and function alone, animal studies are often used to identify pregnancy-related or pediatric toxicity. However, animal studies are limited in their ability to predict human maternal, fetal, or pediatric toxicity, both because of considerable variations in teratogenic and other toxic effects (even among various nonhuman mammalian species) and the usual absence of concordance between effects in animals and humans [54].

Because of the typical exclusions of pregnant women and children from clinical trials, most information regarding the benefit/risk profile of drugs in these populations is collected after a drug's initial approval. In practice, women take drugs while pregnant either intentionally or unintentionally – intentionally because some conditions require treatment during pregnancy, and unintentionally because a large proportion of pregnancies are unplanned [55]. Embryonic exposure to medications at the most vulnerable period of development can, thus, occur before a pregnancy is detected. Similarly, many children are prescribed drugs approved for analogous adult conditions under the potentially wrong assumption of analogous efficacy and safety. Postmarketing data from case reports and case series can offer clues, but except for selected drugs with obvious and dramatic toxicity (e.g., thalidomide), they do not provide conclusive

evidence regarding causality. Selected reporting of exposed cases can lead to false alarms or exaggerated effects, while underreporting may keep other important effects unrecognized [56].

Therefore, postapproval controlled observational studies provide the primary mechanism for identifying potential teratogenic and toxic effects in pregnant and pediatric populations. More generally, pharmacoepidemiologic studies can contribute evidence on the comparative safety and effectiveness of therapeutic strategies during pregnancy and childhood to inform clinical decision-making and policies, and ultimately improve health outcomes in pregnant women and children of all ages. The newer analytic techniques developed to help mitigate confounding in studies of unintended and perhaps intended consequences should prove useful in the context of pregnant and pediatric populations as well (see Chapters 33 and 43).

Methodological Problems to Be Solved by Pharmacoepidemiologic Research

Defining the Population

Pregnant Women

As in any study, it is crucial to specify the target population, inclusion and exclusion criteria, and start and end of follow-up. To prevent biases, including immortal time bias (see Chapter 43), the study design should ideally align start of follow-up, specification of eligibility, and treatment assignment [57]. For example, to assess whether influenza vaccination during the first trimester triggers spontaneous abortion, the follow-up should start at the time of vaccination for the exposed, and at the same gestational time for the unexposed. The peculiarities of pregnancy research start when defining the target population, since the unit of observation may be the mother, the pregnancy (sibling clusters within mother), or the fetus (multifetal clusters within pregnancy). Sometimes twinning or parity is an outcome of interest itself; when it is not the outcome, some studies exclude multiples or select one pregnancy per woman to simplify the analyses, while others account for the correlation within pregnancy and within the mother [58]. In certain cases of treatments with epigenetic or germline effects (e.g., diethylstilbestrol [1,2]), the time of first relevant exposure might occur in grandparents or earlier generations, and relevant outcomes might similarly affect grandchildren or their offspring. Practically, however, most studies will begin at or after conception and continue through the end of pregnancy for maternal outcomes, and through childhood for outcomes of offspring. Given the unplanned and unrecognized nature of many pregnancies, identification of pregnancy at conception and its earlier stages is challenging in routinely collected data. Similarly, given the variability of exposures and risks across stages of pregnancy, identification and consideration of gestational age can be important for many research questions.

Children

The age cutoff for pediatric populations varies across countries, organizations, and agencies, including ages under 17 (US FDA Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research) [59] and under 18 (EMA [50] and the US National Institutes of Health [60]). Biologically, any such cutoff can be misleading; some individuals (usually boys) continue to grow after age 18, and cerebral white matter (along with executive function) continues to mature into one's 20s [61]. For these and other reasons, the American Academy of Pediatrics has discouraged the use of an arbitrary age limit to define pediatrics [62]. Irrespective of how we define pediatrics, age-related developmental changes and the heterogeneity of the pediatric population require consideration of study-specific subgroups in pediatric research, such as infants (premature

and term), children (preschool and school age), and adolescents. As with definitions of pediatrics, the age cutoffs of these subgroups vary across regulatory bodies (US FDA [59], US Census [63], International Conference on Harmonisation [64], EMA [65]). The number and ranges of age subgroups applied to pediatric studies vary even more widely, depending on the study question [66]. Studies that combine children of all ages or combine children with adults, without consideration of differential effects by age, could miss an important source of effect modification and be uninterpretable.

Sample Size Requirements and Challenges

An important practical hurdle when studying the effects of medications in pregnancy and in children is the ability to attain adequate sample sizes because, more often than not, exposures are uncommon, and diseases and outcomes of interest are also uncommon. The importance of considering subgroups and effect modification by stage of pregnancy or pediatric age may increase the sample size needed.

While use of many medications in pregnancy and in children has increased over time (see "Treatment Responses and Patterns"), the proportion of pregnant women or children using a specific medication is still relatively low for most medications [17,21,67]. Likewise, many maternal, fetal, and pediatric adverse outcomes of interest are uncommon. An outcome of particular interest in studies of pregnant women is congenital malformations, which are diagnosed in 2–3% of liveborn infants [68,69]. The prevalence of specific malformations ranges from about 1 per 1000 live births (e.g., oral clefts [70]) to less than 1 per 10000 (e.g., biliary atresia [71]). Similarly, serious outcomes in pediatric populations, such as anaphylaxis, upper gastrointestinal bleeding, diabetes mellitus, or malignancy, are generally rare [72–74]. In the absence of large effects, the study size needed to study the association between less common exposures and outcomes rapidly becomes prohibitive (see Chapter 4 on sample size) [75–77]. To compensate for lack of statistical power, researchers often revert to grouping ("lumping") exposures or outcomes, a potentially problematic solution (see "Exposure Ascertainment, Timing, and Misclassification" and "Outcome Definition and Ascertainment").

A challenge of examining multiple rare exposures in association with multiple rare outcomes is the risk of chance findings in the context of multiple comparisons. All tested results, including the negative and nonsignificant ones, should be reported within studies. Selective interpretation, reporting, or publication of only positive or statistically significant results (e.g., reporting of associations with at least five exposed cases) would lead to publication bias. New findings should be replicated, and the tenability of alternative explanations, including both chance and bias, should be examined before we conclude there is a presence or absence of causal effects [78,79].

Exposure Ascertainment, Timing, and Misclassification

The main goal in ascertaining drug use in any pharmacoepidemiologic study is to minimize the misclassification of the exposure. As a general principle, we strive to define exposure during the etiologically relevant window with high specificity. In the tradeoff between sensitivity and specificity, we prioritize specificity in order to reduce the likelihood that a sizable proportion of women or children classified as exposed are, in fact, unexposed (i.e., false positives). In the context of uncommon exposures, this likelihood is quite high when using an exposure definition with low specificity. Misclassifying unexposed women or children as exposed would dilute the association if one truly exists. There are many approaches to ascertain and define exposure that can be applied to pregnant and pediatric populations (see Chapter 3). Nonetheless, research in pregnancy and children

presents important challenges in defining such exposures accurately.

Pregnancy and childhood are highly dynamic times in medication management, including use of OTC medications [6,80] (see "Changes in Treatment Responses and Patterns"). These challenges are compounded by the frequent lack of a clear definition of the etiologically relevant window and the potential brevity of this window. Inclusion of exposures outside the relevant window generally leads to exposure misclassification and bias toward the null. For many severe birth anomalies, the period of interest is the first trimester, but a narrower window is appropriate for specific anomalies (e.g., the first few weeks after conception for neural tube defects). Given the rapidity of physiologic changes in young children, etiologically relevant windows may be similarly short in duration (weeks to months long) for infants, particularly premature infants, and are likely longer for older children. The etiologically relevant window may be unclear if the pathophysiology of the outcome is poorly understood (e.g., autism spectrum disorder, cognitive delays) or multifactorial (e.g., preeclampsia, childhood obesity), or if the mechanism by which the drug confers excess risk is not known. In these circumstances, it is important to explore different, potentially relevant windows, always acknowledging the exposure windows explored.

To ascertain exposures, we can either rely on secondary data (such as medication dispensings) or primary data collection (such as interviews). Users of secondary data sources should be aware of the disconnect between prescribed, dispensed, and consumed medications and work with highly specific definitions of exposure (see Chapters 12–13). Studies with primary data collection rely on maternal, parental, or adolescent interviews for drug exposure information, which is often the only feasible way to obtain information on OTC drug use and verify the intake of medications. This approach raises concerns about the overall accuracy of recall. Moreover, researchers often conduct such interviews retrospectively after the outcome of interest has occurred (e.g., birth anomalies, pediatric cancer), raising concern about recall bias or differential misclassification of exposure. In theory, the birth of a malformed child or a severe pediatric condition may affect recall of prior, remote exposures (e.g., during pregnancy or infancy). More complete exposure recall among mothers of cases would create a false association between the drug and the birth anomaly or pediatric condition, or overestimate an association if it exists [81–83]. One approach to reducing this bias is improving accuracy by using well-designed interviews with highly structured questions to maximize recall and minimize errors in exposure assessment [84]. Some studies of birth anomalies have used infants with other malformations as controls. Assuming that recall is unlikely to be outcome specific, using mothers of malformed infants as controls helps ensure that reporting accuracy is comparable among mothers of cases and controls [85]. When interviews are used to measure exposure, one should also consider the potential for misclassification due to social desirability bias, for example in studies of opioid use during pregnancy [86] or adolescence [87].

Pregnant Women

Not all pregnancies are 40 weeks in duration, and many outcomes of interest are associated with shorter gestational length (e.g., preterm delivery, preeclampsia, stillbirth). In those instances, one must avoid defining the exposure window in a way that creates *differential opportunity for exposure* in affected and unaffected pregnancies (e.g., exposure during the third trimester), as this will bias the association measure [88,89]. A few different strategies are available to avoid this bias, including (i) ending the exposure window prior to the start of follow-up (e.g., at 22–24 weeks of gestation, the earliest time for viable birth); (ii) defining exposure in a fixed look-back window (e.g., 30 days) from time of delivery; and (iii) using a

time-varying exposure definition (e.g., with a Cox model). Moreover, one outcome affects the opportunity for other outcomes in pregnancy (e.g., competing risk between prematurity and preeclampsia). Depending on the specific question of interest, one may consider using fetuses-at-risk approaches to assess risks that vary with gestational age at birth [90–92] or other methods to account for competing risks (e.g., fetal death) [93,94].

A methodologic concern in pregnant populations related to the grouping of medications is known as the fallacy of "class action" teratogenesis. In reality, members of a given drug class do not necessarily have the same teratogenic or nonteratogenic potential. For instance, we may not know whether teratogenesis results from the chemical structure *common* to a medication class or from the part of the structure that *differentiates* one class member from another. For example, two glutarimides, thalidomide (phthalimidoglutarimide) and glutethimide (phenylglutarimide), are both sedative-hypnotics. Despite these structural and clinical similarities, thalidomide was identified as a high-risk teratogen, and glutethimide was not. Thus, we cannot assume that if one drug is a high-risk teratogen, all other members of its class will share that effect. Conversely, we cannot assume that the safety of one drug in a given class connotes the safety of another. There are, however, some exceptions in which we can hypothesize a potential teratogenic effect based on the existing biologic evidence (e.g., folic acid antagonists and other antimetabolites, which may exert toxic effects through impairment of growing cell division) [95].

Children

Because pediatric dosing is frequently weight based, studies of dose effects in children may rely on accurate determination of weight. Some automated data sources, including administrative claims and some electronic heath record (EHR) databases (see Chapters 12–13), may lack reliable or any data on children's weights. One potential solution is to estimate pediatric dosage by imputing weight based on age and sex, using the appropriate pediatric growth charts (e.g., national, World Health Organization [WHO]), and creating broad weight-based dose categories based on quantiles or prespecified cutoffs [96]. This approach allows for examination of the population-level effects of dose in children whose weight distribution is expected to resemble that of the source population. This assumption may be false when studying conditions that affect pediatric weight, such as malnutrition or other chronic diseases, for whom additional information should be incorporated into any systematic imputation of weight. Additionally, imputed weight-based dosage becomes less valid for more granular estimations of dose, which are more susceptible to misclassification at the individual level. Of note is that weight is not a universal determinant of drug dosing, which may also vary based on other (potentially unmeasured) growth measurements (e.g., body surface area) or pharmacokinetic/pharmacodynamic factors [97,98].

Outcome Definition and Ascertainment

To obtain valid information on outcomes in pregnant and pediatric populations, some studies have direct access to patients and families and use expert clinicians (e.g., teratologists or other pediatric specialists, pathologists) to adjudicate all outcomes. Other studies using administrative claims or EHR data may use restrictive algorithms or validate outcomes using medical records (see also Chapter 37 on validation) [99–101]. It should be noted that birth certificates can be inaccurate records of birth anomalies and are not recommended as gold standards [102].

Replication of a known association (i.e., positive control) can provide reassurance of valid outcome ascertainment. For example, as a positive control, one could replicate the association

between pregestational diabetes (or insulin as a proxy) and major malformations.

Differential misclassification of outcomes with respect to exposure leads to additional analytic challenges. Diagnostic bias may occur if exposed children or mothers-to-be receive more testing (e.g., because of suspicion of teratogenic or other drug-induced effects), resulting in more complete diagnosis or overdiagnosis of subclinical conditions (e.g., minor anomalies or viral illnesses).

Evaluation of long-term outcomes following prenatal or early life exposures, such as developmental or neuropsychiatric disorders, is particularly challenging. First, identification of certain long-term outcomes requires large longitudinal studies with information ranging from less frequent, clinically relevant conditions (e.g., psychotic disorders, malignancies) to more frequent outcomes with less readily available measures (e.g., cognitive impairment based on formal testing, scholastic performance, or educational attainment) or that do not consistently come to medical attention (e.g., autism spectrum disorders). Practically, such studies may be difficult to conduct in settings with high turnover (e.g., local EHR or US administrative claims databases), where relatively few children can be followed continuously for many years after birth. Transition of care from pediatric to adult health professionals can also lead to loss to follow-up because of difficulties in finding new clinicians or transfers to different health systems or insurance. Methods for dealing with censoring should be applied, but loss to follow-up could compromise statistical power and lead to bias if dropout is not random and censoring is informative.

Second, evaluations of the effects of prenatal or early life drug exposure on child development face multiple potential sources of confounding, such as shared home and family environment, genetic predisposition, and the impact of maternal or childhood illness on parenting and healthcare utilization (see "Confounding"). Adjustment for time-varying confounders affected by prenatal or early life exposures (e.g., postpartum maternal mental health) can be problematic, since time-varying factors (e.g., maternal anxiety postpartum) might be affected by prenatal exposures (e.g., maternal use of antidepressants) and affect the outcome of interest (e.g., infant mental health) [103,104]. In this situation, adjustment for such covariates would violate the principle of not adjusting for factors on the causal pathway between exposure and outcome; but not adjusting would violate the principle of adjusting for confounders (e.g., if maternal anxiety postpartum is the only measure of maternal mental health available).

Pregnant Women

Drugs can induce a wide range of reproductive outcomes, including hampering fertility and interacting with contraception. Drugs given to pregnant women may also lead to adverse obstetric outcomes, perinatal complications, and even neurodevelopmental delays in offspring later in life. Effective drugs can also reduce adverse outcomes induced by the underlying condition (e.g., reduction in obstetric complications by improving glycemic control in women with diabetes). Among outcomes of interest for pregnant women, teratogenesis has received special attention for being a rare but dramatic event. In recent years, however, pregnancy researchers and regulators have expanded their attention to include fetal losses and long-term consequences in the child.

Researchers sometimes lump together various fetal malformations, partially for conservation of statistical power [105]. However, given the etiologic heterogeneity of malformations, the combination of multiple malformations into a single outcome may lack a sound embryologic basis, even when such outcomes are classified by organ system (e.g., cardiovascular) [106]. A more appropriate approach may be to create categories that reflect the embryologic tissue of origin (e.g., neural crest [107]) or teratogenic mechanism (e.g., disruption of the embryonic vasculature [108]), when known [109].

Depending on the outcome definition, the risk of malformations overall can range from 1% to over 10% [110]. Common exclusion criteria when evaluating teratogenicity include chromosomal anomalies, known genetic disorders, minor anomalies, birth marks, positional deformations, subclinical anatomic findings by ultrasound, and complications of prematurity [111]. In some circumstances, confirmation may require specific procedures (e.g., an echocardiogram to measure the size of a septal defect) or follow-up (e.g., patent foramen ovale persistent six weeks after birth). Another source of heterogeneity comes from the period of observation: some studies may include prenatal diagnoses and over one year of follow-up after birth; other studies may focus on an outcome ascertainment window of days around delivery [112]. Therefore, estimates of the risk of malformations from one study might not be comparable with estimates from birth anomaly surveillance systems or other studies with different inclusion and exclusion criteria [113]. Consequently, investigators should strongly consider including an internal reference group with consistent data collection and applying the same criteria to measure outcomes [114].

To estimate the risk of fetal losses, *ad hoc* pregnancy cohorts need to enroll women soon after conception to consider miscarriage, and they require large sample sizes to consider stillbirths (expected frequency 6 per 1000 in the US [115]). When comparing treatments, researchers must ensure that all groups have comparable gestational age at enrollment and that differential enrollment time (left truncation) is handled correctly in the analysis [116].

Children

Given the many clinical and epidemiologic differences between children and adults (see "Unique Biology and Epidemiology"), pediatric researchers should be careful when applying validated definitions from adult populations to children. For example, depression (like many chronic diseases) is rarer in children than in adults, thus reducing the positive predictive value of a given algorithm, assuming equal sensitivity and specificity. One should also not assume that conditions with the same name or code represent the same disease in children as in adults, for instance neutropenia, since children are much more likely to have cyclic or chronic neutropenia [117]. As another example, compared with adults, children with diabetes mellitus are more likely to have type 1 diabetes, for which more specific diagnostic codes exist (although they may not reliably be used). Furthermore, type 1 diabetes is quite distinct from type 2 diabetes in pathophysiology, clinical presentation (e.g., thin children who are more likely to present with diabetes ketoacidosis than adults with incident type 2 diabetes), and treatment (i.e., exclusively with insulin). Even conditions that are managed more similarly in pediatric and adult populations (e.g., juvenile idiopathic arthritis versus rheumatoid arthritis, bipolar disorder) may have differences in prevalence, clinical presentation, coding, or outcomes that necessitate independent validation [118]. Thus, depending on the study question and setting, an outcome validated in adults should not be presumed to be valid in children, and vice versa.

As already mentioned, milestones in growth and development can serve as important outcomes in pediatric studies. For diseases that impair growth or delay puberty, for instance, by disrupting nutrition, absorption, or metabolism (e.g., thyroid disease, inflammatory bowel disease, juvenile idiopathic arthritis), improved or normalized growth and development can be measures of drug effectiveness [119–121]. Additionally, growth and development can also be important markers of safety (e.g., slowed vertical growth with glucocorticoids [122], impaired weight gain with stimulants [123], excessive weight gain with antibiotics [124]).

Confounding

Because of the lack of randomization in observational studies, there is no guarantee of balance in characteristics between comparison groups. Confounding by indication can occur if children or pregnant women who receive a drug are more likely to have an underlying condition, or a more severe or active form of that condition, which is associated with the risk of the outcome. Available approaches to minimize confounding by indication include (i) adjusting for the presence of the underlying condition; (ii) restricting both the treated and reference group to individuals with a recorded diagnosis for the underlying condition; (iii) restricting to pregnant women or children with the indications and using an active reference group; (iv) comparing continuers with discontinuers of the medication of interest (more applicable to studies of pregnancy); and (v) sibling discordance study to control for stable family factors (see "Currently Available Solutions, Newer Designs"). Confounding by disease severity poses additional challenges in observational research on medication safety and effectiveness (see also Chapter 33).

Aside from confounding by indication and disease severity, other putative risk factors for the outcome may be associated with exposure. One advantage of using large EHR or claims databases for research in pregnant women and children is that the researcher can exploit the richness of the data to identify and control for a large number of potential confounders or proxies for these variables. Given that outcomes tend to be rare, and the set of prespecified potential confounding variables may be large, the use of data-reduction techniques such as propensity scores can help avoid problems with model overfitting (see Chapter 43).

Many secondary data sources, including EHR and claims data, do not have robust or complete information on certain potential confounders (e.g., socioeconomic status; smoking, alcohol use, or recreational drug use in adolescents, pregnant women, or caregivers; test results such as prenatal testing) and lack information on OTC medications (e.g., multivitamins, folate, analgesics and antipyretics; see Chapters 11–14). Several approaches can mitigate potential residual confounding from unmeasured factors, including the use of high-dimensional propensity score analyses (if unmeasured confounders are correlated with measured ones) [125] or external adjustment for unmeasured confounders [126,127]. After utilizing all approaches to minimize confounding during the stages of study design and analysis, one can conduct quantitative bias analysis, a sensitivity analysis that incorporates the level of uncertainty in estimates based on suppositions about unmeasured confounders (see Chapter 43) [128]. Another approach that provides reassurance about control for confounding is replication of a null association (i.e., negative control). For example, one could document a null association between paternal use of the drug at the time of the pregnancy and the risk of fetal growth restriction.

Pregnant Women

Although many treatment indications are not traditional risk factors for adverse outcomes of pregnancy, certain indications may be confounded due to strong associations with other conditions or behaviors that are risk factors for the outcome. For example, women treated with antidepressants (including women with depression and/or anxiety) may also be more likely to smoke, use substances, eat poorly, and have chronic conditions such as diabetes, hypertension, and obesity – all factors that predispose to congenital heart anomalies [129]. In addition, women with anxiety may utilize more healthcare resources, including fetal or early-life testing (e.g., echocardiography), than their unaffected counterparts. Hence, anxious women are more likely to have infants diagnosed with mild

cardiac malformations that might have gone clinically undetected in children of other women, such as small muscular ventricular septal defects, which often remain subclinical before resolving spontaneously [130]. Failure to account for such sources of confounding when studying the safety of psychotropic medications might bias the results.

Children

Certain sources of confounding may play a large role in pediatric studies, including second-hand smoke exposure, parental income and occupation (e.g., measures of socioeconomic status or environmental exposure), parenting styles (including behaviors related to healthcare utilization; e.g., requests for testing or antibiotics, etc.), early childhood feeding practices (e.g., breastmilk or formula), medical conditions of parents and siblings (and more fundamentally, genetics), and vaccination status [131]. Birth weight and gestational age at birth can be particularly important to understanding the indications and effects of treatment in the earliest days and months of life, especially in premature neonates and infants. Many of these variables can be difficult to measure in large data sources, particularly in administrative claims data.

Analyses of pediatric populations often need to adjust for age or stratify by age group with greater precision than in adult studies. Whereas 10-year age groupings in studies of adults may be highly appropriate, there are dramatic differences between a 4-year-old and 14-year-old child, and even between a preterm infant, a term infant, and a 2-year-old (see "Unique Biology and Epidemiology" and "Defining the Population"). Thus, representation of age and age groups must consider the pediatric population and clinical question of interest. Date of birth, a variable that is unavailable in some de-identified data sources, can be an especially important consideration in studies of newborns, infants, and other young children. For example, date of birth can relate to seasonal events (e.g., viral and bacterial infections) and other temporal changes (e.g., changes in vaccination practices) in ways that are critical to interpreting associations between key drug exposures (e.g., antibiotics) and outcomes (e.g., asthma) [132]. Date of birth and birth cohort could, thus, be considered as a confounder, effect modifier, or even instrumental variable. At the earliest extreme, neonatal studies may require information about the hour of birth to evaluate outcomes that occur within several hours or days of birth.

In addition to considerations of age, states of growth and development could also be sources of confounding. For instance, malnourished or obese children may have increased risks for receiving treatment (e.g., antibiotics) and experiencing an outcome of interest (e.g., treatment-refractory infection). Growth and development can also be a source of effect modification. For instance, the robust skeletal turnover and repair in growing children may protect them from adverse effects of drugs that damage bone in fully grown adolescents and adults [96,133,134].

Selection Bias

When selection into or retention in the study is directly or indirectly affected by the exposure and the outcome, selection bias may occur and distort the estimate of risk [135].

Pregnant Women

In nonpregnant populations, the importance of including new users to avoid a selection of exposed subjects nonsusceptible to potential adverse effects is well recognized [136]. In pregnancy, there are also situations in which considering prevalent users complicates the interpretation of results [136]. For example, in evaluating the effect of certain antipsychotic medications on the risk of gestational diabetes, inclusion of women on the drug at conception

who do not have diabetes at baseline might exclude patients susceptible to the cardiometabolic effects of the antipsychotics; those pregnant women entering the cohort, therefore, might not be susceptible to developing diabetes with drug exposure, biasing results to the null [137].

In clinical practice, recommendations for interventions in early pregnancy often affect not only women actively trying to become pregnant, but also any women of childbearing age, given that a large proportion of pregnancies are unplanned and there is no time to switch treatments before organogenesis and after pregnancy is recognized [55]. Similarly, in research, the ideal pregnancy cohort begins at or before conception, if not at the time of first exposure [138]. This situation occurs rarely in reality. More often studies enroll women after confirmation of a pregnancy. This form of left truncation may underestimate the risk of early pregnancy events (e.g., miscarriages). Differences in gestational age at enrollment between exposed and reference groups could lead to biased relative risk estimates if the risk varies with gestational age. For example, consider a study investigating how a vaccine given during the first trimester affects the risk of miscarriage, which decreases after week 10 of gestation. A design that erroneously started follow-up for exposed individuals after vaccination and for unexposed individuals after a positive pregnancy test would introduce selection bias. To prevent this bias when evaluating the risk of early miscarriages, the safest approach is to enroll subjects as soon as possible after conception, and to start follow-up of exposed and unexposed at comparable gestational ages.

Selection bias may also occur if the outcome affects enrollment differently in exposed and reference groups [116]. For example, concerning results from prenatal screening may motivate women on specific medications to enroll in a pregnancy registry. This scenario would lead to overestimation of risks among exposed women [116]. On the other hand, a study could preferentially select low-risk pregnancies if the investigators declined to enroll women with concerning prenatal screening results, or if women were less likely to enroll after a therapeutic abortion. To prevent this bias when evaluating the risk of birth anomalies or other complications of pregnancy, the safest approach is to enroll subjects before the risk of the outcome is known; that is, before the completion of informative screening tests.

An issue unique to the study of birth anomalies is the possibility of pregnancy losses, whether spontaneous or induced. Many studies are restricted to live births, and studies that include pregnancy losses almost always lack information from pathology reports on the presence of structural anomalies. Moreover, some malformations become detectable at early stages of pregnancy, and some women decide to terminate these pregnancies. Studies of liveborn infants, thus, underestimate the risk of lethal and prenatally detectable anomalies. The proportion of fetuses identified to have birth anomalies after termination varies by type of malformation, ranging from under 5% for oral clefts to more than 40% for neural tube defects [139,140]. Bias could occur in instances where exposed and reference groups had different proportions of terminations of affected fetuses. For example, if women exposed to a putative teratogen were more likely to terminate a pregnancy with a malformed fetus, and this situation was not captured within the study, the relative risk estimate would be biased toward the null. Although the likelihood of terminating a pregnancy based on an adverse prenatal diagnosis might not be affected by drug utilization [141], this assumption is usually untestable. Sensitivity analyses can be used to assess the uncertainty around the relative risk estimates due to selection bias, together with misclassification and confounding [142]. These methods do not substitute for a valid and carefully conducted study design, but can produce a plausible range of estimates under realistic assumptions [143].

Sometimes, studies include pregnancy losses without fetal autopsy information on structural malformations. These studies can include all pregnancies in the denominator of risk estimates, but still face the challenge of outcome classification in the numerator. Missing data after spontaneous abortions or terminated pregnancies could lead to false negatives. If this outcome misclassification preferentially affected exposed women, this ascertainment bias would underestimate the relative risk of the drug. This situation can be considered a problem of right truncation, where follow-up ends with unknown outcomes; or of competing risk where fetal demise competes with the risk of being born with a malformation. Similar competing risk circumstances arise at the end of pregnancy, where pregnancy ends before an outcome can occur (e.g., competing risks between preterm delivery and the development of preeclampsia).

Case–control studies confront an additional type of selection bias resulting from inappropriate control selection. In some studies of birth anomalies, controls comprise infants with malformations other than the ones of interest affecting cases [144]. This approach is valid under the assumption that teratogens do not increase the risk of all malformations, including the specific malformation among controls [145,146]. The goal of using a malformed control group is to reduce the opportunity for differential recall of exposure between mothers of cases and controls. Whether malformed or not, controls should be sampled from the same population that gave rise to the cases. In multicenter case–control studies that identify cases from a large number of hospitals, controls should be identified from the same hospital catchment area [147], and the analyses should account for matching of cases and controls by center [148].

Lastly, one can introduce selection bias during the analysis by adjusting for variables that share common causes with the outcome or are affected by it. Thus, adjusting for these variables (e.g., low birth weight) will not reduce confounding and can introduce selection bias [135,149,150]. For example, adjustment for low birth weight is unwarranted when the analytic goal is to estimate the overall effect of prenatal variables, such as maternal drug use, on infant mortality, or when the goal is to estimate the direct effect but there is an unmeasured common cause of low birth weight and mortality [150]. Knowledge of the causal structure is a prerequisite to accurately labeling a variable as a confounder. In studies of pregnancy, events that occur after organogenesis cannot cause structural birth anomalies arising during the first trimester and, therefore, are not confounders. Nonetheless, they could be proxies for strong unmeasured confounders, in which case investigators may want to use them for statistical adjustment to reduce confounding.

Children

Large longitudinal cohorts and registries can be valuable settings to study the effects of drugs in pediatric populations (see "Prospective Cohorts" and "Registries"). However, differential entry or loss to follow-up because of factors associated with the outcome (e.g., disease severity, psychosocial risk) can introduce selection bias [151,152]. Selection bias may also affect research of other special pediatric populations at high risk of poor outcomes. Research in very premature infants is often conducted within highly specialized referral neonatal intensive care units. Studies in which referral patterns and, thus, study inclusion relate to the risk of exposure (e.g., indomethacin) and outcomes (e.g., intraventricular hemorrhage, mortality) may result in selection bias [153,154].

Another highly vulnerable pediatric population is children in foster care, who are subject to early trauma and inconsistent preventive medical care, leading to psychosocial disruption and potential overtreatment (e.g., with psychotropic drugs) [155,156]. One may use Medicaid claims data to study this vulnerable population, but Medicaid eligibility codes misclassify children who are in and not in foster care; correctly classified children may be more likely to have higher

levels of healthcare utilization and case management support [157,158]. As a result, studies on the effects of certain drugs in foster children may be biased in samples in which excluded children (e.g., those with low levels of support) have different risks of both exposure (e.g., to antipsychotics) and the outcome of interest (e.g., diabetes mellitus). Studies may reduce the risk for selection bias by accounting for these coding practices and utilizing supplemental data to better classify foster care status.

Currently Available Solutions

In this section we review the main pharmacoepidemiologic designs used for pregnant and pediatric populations to quantify the risk/benefit profile of medication exposure during pregnancy, infancy, or childhood, and the solutions they offer to common forms of bias.

Prospective Cohorts

Pregnant Women

Prospective inception (or follow-up) cohorts of pregnant women have the advantage of identifying drug exposure before the adverse outcomes are recognized. This approach involves the identification of a population of women to be followed as soon as possible after conception or at the stage of pregnancy planning, to allow the evaluation of early pregnancy events. Such studies should periodically collect information on demographic characteristics, exposures, and potential confounders, as well as formally evaluate offspring at birth (or fetal death) and ideally throughout childhood. Unless regular pregnancy tests are conducted, studies will not detect fetal losses that occur before pregnancy is recognized. Because birth anomalies and stillbirths are rare outcomes, and because usually a small fraction of pregnant women will take particular drugs, inception cohorts need to be large. Even exceptionally valuable research resources such as the Danish National Birth Cohort [159] or the Norwegian Mother and Child Cohort Study (MoBa) [160], with over 90 000 pregnancies, were too small to examine the risks of specific birth anomalies related to specific drugs taken in pregnancy.

Children

Prospective cohorts may also permit evaluation of a variety of pediatric exposures and outcomes of interest, while allowing for detailed adjustment for confounders. Birth inception cohorts of premature infants and children are particularly useful to study the link between early life treatments and subsequent outcomes [161,162]. Birth inception cohorts have been used to study term infants as well, for example the link between probiotic use in the first month of life and subsequent development of type 1 diabetes-related autoantibodies among genetically predisposed children [163]. Disease-focused prospective cohorts may also be useful for pharmacoepidemiologic research on children with rare diseases, such as chronic kidney disease [164], congenital heart disease [165], inflammatory bowel disease [166], juvenile idiopathic arthritis [167], psychotic disorders [168], and venous thromboembolism [169]. Disease-inception cohorts of treatment-naive children may be less susceptible to selection bias than other prospective cohorts that voluntarily enroll children who have already received treatment for their condition (see "Selection bias"). Like other observational settings for pediatric research, prospective cohorts are subject to a variety of methodologic limitations, including confounding and limited sample size (see "Methodological Problems").

Registries

Pregnant Women

For new or infrequently used drugs, it is more efficient to assemble cohorts of women exposed to a drug of interest in pregnancy and follow them to determine outcomes (known as exposure pregnancy registries) [170,171]. The pri-

mary objective of many exposure registries is to assess the relative risk of major congenital malformations in the offspring. However, registries can evaluate multiple maternal, obstetric, fetal, and infant outcomes [171]. These efforts are critically important for the detection of major adverse effects (e.g., isotretinoin teratogenicity) [25] that affect large proportions of exposed fetuses. However, the small size of most registries (usually with no more than a few hundred women) prohibits the identification or disproof of small or moderate effects involving rare outcomes [172].

To avoid selection bias, women should be enrolled into a pregnancy registry before the pregnancy outcome is known. As in other studies based on volunteers, self-referral often results in a nonrepresentative population. Similarly, the participants who complete follow-up may be even less representative. Selective inclusion or follow-up may affect the generalizability of absolute risk estimates, and may bias the relative risk if selection or retention is related to both the exposure and the outcome.

In any pharmacoepidemiologic study, it is important to include comparable reference groups. Many registries only enroll exposed women and compare the incidence of malformations with estimates from surveillance systems. This approach raises concerns about the comparability of control groups, including demographic and clinical characteristics as well as the ways in which outcomes are defined, detected, and validated. Other registries compare the observed risks in exposed women to those in pregnancies without exposure or with exposure to unrelated (e.g., presumed nonteratogenic) drugs. These approaches rarely consider confounding by indication. As discussed earlier (see "Confounding"), one should compare women exposed to a drug of interest with other women who have similar indications, whether untreated or treated with alternative drugs [173]. Therefore, when feasible, it is preferable to establish multidrug pregnancy registries that allow comparisons among drugs from the same class or indication. More recently, both private and government initiatives have launched pregnancy registries using health apps for direct data collection from large populations of pregnant women (e.g., PregSource®). The validity of these data for drug safety research has not yet been established.

Children

Like other prospective disease cohorts, pediatric registries for children with rare chronic diseases can provide rich information about exposures, outcomes, and confounders [174–176]. Rare pediatric disease registries may also collect biologic specimens that facilitate molecular pharmacoepidemiology [177–179]. Some pediatric registries are population based and comprehensive in their ascertainment of affected children and, therefore, can minimize selection bias and yield generalizable knowledge [180]. Other registries rely on participants or families to voluntarily enroll and may be, thus, subject to selection bias or reduced external validity [181,182].

Retrospective Cohorts and Nested Case–Control Studies within Automated Healthcare Databases

Population-based automated healthcare databases, including national registries (e.g., Nordic registers), administrative claims databases (e.g., Medicaid), and EHR databases (e.g., Clinical Practice Research Datalink®, CPRD®), are frequently used sources of information for pharmacoepidemiologic studies (see Chapters 11–14) [183]. These databases allow researchers to conduct large-scale observational postmarketing studies on multiple rare or long-term consequences of drug use [184]. While these resources have their own strengths and limitations, they all offer detailed, longitudinal records of healthcare utilization, diagnoses, procedures, and drug data (prescriptions, dispensings, or

both) across a range of healthcare settings. The clinical care represented in these databases reflects the real world, and study populations may include minorities and other marginalized populations that are often excluded from volunteer-based studies. Although the cost and time of working with large administrative and EHR datasets can be high, this approach is usually less costly and has greater breadth than primary data collection. The size and generalizability of population-based automated healthcare databases can make them excellent settings for studies of rare exposures or outcomes in pregnant women and children. However, some automated healthcare databases have substantial limitations due to lack of child–mother linkages or routinely collected data on gestational age, birth weight, maternal obesity, smoking, or use of nonprescription drugs [185].

Pregnant Women

Linkage of children's records with those of their mothers is relatively comprehensive in Nordic countries, which have nationwide registries with national identification numbers that help link family members. Linkage between maternal and children's records in US healthcare databases, integrated health systems, and most EHR databases is also feasible [186–189]. US administrative claims databases lack information on gestational age, which needs to be estimated based on codes for preterm, term, or post-term delivery [190,191].

When exposure to the specific drug of interest involves a small fraction of the pregnant population, even these large cohorts are constrained in their information. In this scenario, multisite collaborations offer a solution. The Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP) is a collaboration among government, academic research centers, and healthcare organizations to combine large administrative databases, linking data together from mothers, babies, and birth certificates [189]. This work has been continued in the Sentinel network [192] and is an important tool to study outpatient dispensing of medications during pregnancy and a number of validated pregnancy outcomes. More recently, international collaborations have allowed the identification of exposed pregnancy cohorts nested in multiple large healthcare databases with the goal of conducting not only surveillance but also etiologic research on the safety of medications during pregnancy. For example, the goal of the International Pregnancy Safety Study (InPreSS) consortium is to provide the best available human data on the safety of prescription medications during pregnancy by combining large-scale data from several countries [193]. Currently, it includes nationwide US Medicaid data that capture nearly half of all pregnancies, plus the national registries in the five Nordic countries that capture virtually all pregnancies. Identification of exposed cohorts nested within an international network of population-representative, prospectively collected datasets replicates the traditional exposure pregnancy registries in a cost- and time-efficient manner. This approach also avoids many of the potential biases that face *ad hoc* pregnancy registries [194]. One can also use collaborative research programs to rapidly follow up on safety signals initially identified in a single data source, thus reducing the chance for widespread worry based on concerning early findings that are not later substantiated. Premature dissemination of such false alarms creates challenges for pregnant women and their clinicians trying to make the best treatment decisions.

Children

A variety of automated databases have been used to study drug uses and effects in large pediatric populations, include encounter-based databases [195–198] and national EHR databases [199–201]. EHR databases are more likely to contain relevant pediatric data, such as birth weight, gestational age at birth, and growth measurements (see also Chapter 13). Some

databases that are also used to study pregnant women can be used to assemble retrospective birth inception cohorts that can jointly examine prenatal and postnatal exposures (e.g., antibiotics) in relation to subsequent outcomes [202,203]. As in pregnancy-related research, multinational collaborations of pediatric researchers, such as the Global Research in Paediatrics (GRiP) Network of Excellence and the Task-force in Europe for Drug Development for the Young (TEDDY) Network, have allowed for large-scale studies on drug utilization and effects, including rare outcomes, in children [72,204,205]. Within the US, there are several collaborative networks of pediatric EHRs that can facilitate pharmacoepidemiologic research, including the Comparative Effectiveness Research through Collaborative Electronic Reporting (CER^2) Consortium [206,207] and PEDSnet [208]. One population-representative EHR database in Italy, Pedianet, is composed completely of data from family pediatricians (see Chapter 13).

Several hospital-based databases have been used for pediatric pharmacoepidemiology research (see also Chapter 14). These include the Pediatric Hospital Information System (USA, >40 freestanding children's hospitals), Pediatrix (USA, >350 neonatal intensive care units), and the Vermont Oxford Network (global, >1000 neonatal intensive care units). Inpatient pediatric databases can provide useful settings for studying inpatient drug uses and effects in children with serious diseases, including extreme prematurity, rare diseases, and children with complex chronic conditions [209–213]. A common limitation of all these databases is the lack of generalizability based on the settings of care (predominantly academic pediatric hospitals) and the lack of outpatient data.

For certain pediatric outcomes, linkage of children's records to mothers as well as other family members, including fathers and siblings, may also be as important, because of concerns for confounding from genetic factors or other shared household factors (see "Confounding" and "Newer Designs") [214,215]. Databases with household/family identifiers (e.g., Medicaid, The Health Improvement Network® [THIN®]) or information on biologic relationships (e.g., Swedish registries) facilitate linkage to family members [124,216,217]. Of note is that residence identifiers do not necessarily specify which household members are biologically related, and one might need to make assumptions about relationships based on age and gender (e.g., mother vs. father).

Outside of the home, school is one of the most important places in a child's life and, thus, a unique setting for studies on medications and devices used by children. Pediatric researchers may want to study the impact of medications on outcomes such as attendance, promotion, behavior, and test performance. In some settings, it may be possible to link health or medication data to scholastic or testing information [218,219]. In settings where one cannot easily link clinical data to school records, researchers may need to rely on self-report [220] or direct cognitive assessments [221,222].

Case–Control Studies

Case–control studies identify individuals with the outcome of interest (e.g., a specific birth anomaly or childhood illness) and compare their frequency of exposure to that in a control group without this outcome. This design offers advantages in the evaluation of associations between prenatal or early life exposure to relatively common medications and the risk for rare events [223,224]. Case–control studies have some important limitations, however. They collect information on exposure retrospectively [225,226], rarely have a sufficiently large sample size to evaluate infrequently used medications, and sometimes do not properly select controls from the same source population as the cases. Furthermore, case–control studies typically focus

on only one type of outcome (e.g., major congenital malformations) and can estimate relative risks, but not the absolute risks associated with the drug, unless the study is nested within a defined cohort or otherwise contains additional information.

Pregnant Women

Case–control studies on birth anomalies are often based on interviews, thus allowing the collection of information on actual use of the drug (versus prescription or pharmacy dispensing), nonprescription drugs, and important covariates such as smoking, socioeconomic factors, or obesity, which are often missing from other data sources. Examples include the Slone Epidemiology Center birth defects study [35]; the National Birth Defects Prevention Study [67], which involved a number of state birth anomalies surveillance programs and was coordinated by the US Centers for Disease Control and Prevention; EUROmediCAT, which incorporates data on prenatal drug exposures and birth anomalies from multiple European registries [227]; and the ECLAMC (Latin-American Collaborative Study of Congenital Malformations) network from Latin America [228].

Children

Many pediatric case–control studies are limited to particular outcomes of interest on a study-by-study basis. In contrast, the Canadian Pharmacogenomics Network for Drug Safety uses a systematic case–control approach to study pharmacogenomic factors associated with a variety of serious outcomes in thousands of children. The network's focus is on children with cancer, because of their high burden of exposure to toxic medications. The network collects detailed clinical information and genomic data on children with recognized serious adverse drug effects and matched controls. In the future, this network plans to make its database and analytic tools publicly available to other pediatric researchers.

Newer Designs

Epidemiologists continue to explore more valid and efficient approaches to study pregnant women and children. In specific circumstances, when carefully conducted with clearly stated assumptions and interpretation of estimates, novel designs may bring advantages to the field.

Pregnant Women

To avoid between-person confounding, one might study the risk of birth anomalies using multiple exposure windows within one pregnancy (i.e., self-controlled design) [229] or multiple pregnancies in the same woman with discordant exposure status (i.e., sibling discordance study) [230]. However, the risk of time-varying within-person confounding remains (e.g., indication for the drug at a particular time) [229]. In a self-controlled design, one studies the presence and timing of exposure in people with the outcome of interest (see also Chapter 43 on self-controlled designs). For example, to study whether flu vaccine administration triggers miscarriage, one could compare the frequency of vaccinations during the month before the event with the frequency in a one-month control window three months before the event. However, in the presence of gestational time trends of drug utilization, case-crossover designs do not offer clear advantages over cohort or case–control designs to study birth anomalies [229]. In a sibling discordance study, the comparison is made between siblings born to the same parents but who differ with respect to their pregnancy exposure status (i.e., matching within mother rather than within pregnancy). This design exploits the fact that siblings share stable aspects of family context as well as half their genome, therefore accounting for unmeasured genetic and environmental factors that may be important sources for confounding. Sibling discordance studies are particularly valuable to test associations identified in studies of unrelated individuals where unmeasured confounding is a concern [231].

Children

Self-controlled designs, including case-crossover and self-controlled case series, have also been applied to drug and vaccine safety research in children [232,233]. Self-controlled designs are particularly appropriate for studies of incident, short-term exposures (common in children [21]) and acute outcomes, including rare conditions that occur in pediatric populations. As in research on pregnancy, self-controlled designs in children could be problematic if study periods extend across times of considerable personal change (e.g., first year of life) without adequate adjustment for age and other time-varying confounders [234]. Sibling discordance studies have also been used successfully to validate, and sometimes refute, pediatric treatment–outcome associations first suggested in traditional cohort studies [202,235]. In research on treatment effectiveness, instrumental variables [236–238] and pragmatic trials [239,240] have been applied to pediatric populations.

The Future

Professional organizations and governments have increasingly supported policies and regulations that prioritize research on medicines and devices for pregnant women and children. Large-scale, longitudinal, collaborative research using multiple data sources across multiple countries will become increasingly common and important for generating generalizable, actionable evidence for pregnant and pediatric populations. Collaborative networks and pooled resources, potentially through distributed data models, will enable the conduct of robust research on rare exposures and rare outcomes in pregnant women and children, including studies on specific drugs, drug dosage, and polytherapy in individuals with chronic conditions. Advancements in pregnancy-related and pediatric pharmacoepidemiologic research will require not only larger data sources and collaboration, but also more transparency and sharing of protocols, data, and analytic code and tools. In particular, future efforts must help build capacity, expertise, and infrastructure to conduct pharmacoepidemiologic studies in underserved settings and low- and middle-income countries.

The epidemiology of diseases of pregnancy and childhood continues to evolve, with the rising worldwide prevalence of obesity and associated ailments, emerging infectious diseases affecting pregnant women and children (e.g., Zika virus), increases in survival from life-threatening pediatric diseases (e.g., cystic fibrosis, malignancy, neuromuscular disorders), and many others. These changes in disease epidemiology will necessitate the generation and sharing of new evidence regarding drug safety and effectiveness for new indications in pregnant women and children. There remains a dearth of high-quality research on treatment effectiveness in these populations, in part because of the fundamental challenges of conducting observational effectiveness research (see Chapter 33). New methods on studying effectiveness and controlling for confounding by indication and disease severity will help improve the quality and quantity of effectiveness research for pregnant women and children. More research and better methods are also needed to study drug uses and effects in lactating mothers and breastfed infants. Another population that warrants further attention is pregnant youth under age 18, who represent an ethically challenging and potentially high-risk subgroup that is commonly excluded from pregnancy-related research.

New and growing linkages between complementary types of data – automated databases, registries, patient-generated data, biobanks containing genomic and other -omic data, and others – will open up new frontiers for the discovery and validation of personalized treatment regimens. From a methodologic perspective, linkage between datasets will also facilitate the validation of outcomes and better

control for confounding through techniques such as propensity-score calibration [241]. Because large healthcare databases will not be sufficient or appropriate for all research questions about pregnant and pediatric populations, we should continue to use, improve, and teach field methods for primary data collection (including biospecimens for pharmaco-omic studies) from patients and families. With advancements in technology and expansions in information, clinicians, pregnant women, and families will need enhanced resources to curate and interpret the deluge of available data. Bedside tools for data analyses (e.g., pharmacogenomic risk stratification) as well as shared decision-making will help clinicians, patients, and families understand and discuss the risks and benefits of treating pregnant women and children in the face of clinical uncertainty, and make informed, personal decisions [242–245].

References

1 Hernandez-Diaz S. Iatrogenic legacy from diethylstilbestrol exposure. *Lancet* 2002; **359**(9312): 1081–2.

2 Giusti RM, Iwamoto K, Hatch EE. Diethylstilbestrol revisited: a review of the long-term health effects. *Ann Intern Med* 1995; **122**(10): 778–88.

3 Carapetis JR, McDonald M, Wilson NJ. Acute rheumatic fever. *Lancet* 2005; **366**(9480): 155–68.

4 Goldkind SF, Sahin L, Gallauresi B. Enrolling pregnant women in research: lessons from the H1N1 influenza pandemic. *N Engl J Med* 2010; **362**(24): 2241–3.

5 Mitchell AA, Gilboa SM, Werler MM, Kelley KE, Louik C, Hernández-Díaz S. Medication use during pregnancy, with particular focus on prescription drugs: 1976–2008. *Am J Obstetr Gynecol* 2011; **205**(1): 51, e51, e58.

6 Thorpe PG, Gilboa SM, Hernandez-Diaz S, *et al*. Medications in the first trimester of pregnancy: most common exposures and critical gaps in understanding fetal risk. *Pharmacoepidemiol Drug Saf* 2013; **22**(9): 1013–18.

7 Chambers K. Asthma education and outcomes for women of childbearing age. *Case Manager* 2003; **14**(6): 58–61.

8 Seyberth HW, Kauffman RE. Basics and dynamics of neonatal and pediatric pharmacology. In Seyberth HW, Rane A, Schwab M (eds), *Pediatric Clinical Pharmacology*. New York: Springer, 2011, pp. 3–49.

9 Brouwer KL, Aleksunes LM, Brandys B, *et al*. Human ontogeny of drug transporters: review and recommendations of the Pediatric Transporter Working Group. *Clin Pharmacol Ther* 2015; **98**(3): 266–87.

10 Sadler NC, Nandhikonda P, Webb-Robertson BJ, *et al*. Hepatic cytochrome P450 activity, abundance, and expression throughout human development. *Drug Metab Dispos* 2016; **44**(7): 984–91.

11 Johnson PJ. Neonatal pharmacology: pharmacokinetics. *Neonatal Netw* 2011; **30**(1): 54–61.

12 Marshall JD, Kearns GL. Developmental pharmacodynamics of cyclosporine. *Clin Pharmacol Ther* 1999; **66**(1): 66–75.

13 Takahashi H, Ishikawa S, Nomoto S, *et al*. Developmental changes in pharmacokinetics and pharmacodynamics of warfarin enantiomers in Japanese children. *Clin Pharmacol Ther* 2000; **68**(5): 541–55.

14 Murrin LC, Sanders JD, Bylund DB. Comparison of the maturation of the adrenergic and serotonergic neurotransmitter systems in the brain: implications for differential drug effects on juveniles and adults. *Biochem Pharmacol* 2007; **73**(8): 1225–36.

15 Expert Panel on Therapeutic Products for Infants Children and Youth. Children are not small adults: considering variation in drug response. In *Improving Medicines for Children in Canada*. Ottawa: Council of Canadian Academies, 2015, pp. 49–88.

16 Chai G, Governale L, McMahon AW, Trinidad JP, Staffa J, Murphy D. Trends of outpatient prescription drug utilization in US children, 2002–2010. *Pediatrics* 2012; **130**(1): 23–31.

17 Hales CM, Kit BK, Gu Q, Ogden CL. Trends in prescription medication use among children and adolescents: United States, 1999–2014. *JAMA* 2018; **319**(19): 2009–20.

18 Man KKC, Ip P, Hsia Y, *et al*. ADHD drug prescribing trend is increasing among children and adolescents in Hong Kong. *J Atten Disord* 2017; **21**(14): 1161–8.

19 Bachmann CJ, Wijlaars LP, Kalverdijk LJ, *et al*. Trends in ADHD medication use in children and adolescents in five western countries, 2005–2012. *Eur Neuropsychopharmacol* 2017; **27**(5): 484–93.

20 Zhang T, Smith MA, Camp PG, Shajari S, MacLeod SM, Carleton BC. Prescription drug dispensing profiles for one million children: a population-based analysis. *Eur J Clin Pharmacol* 2013; **69**(3): 581–8.

21 Sturkenboom MC, Verhamme KM, Nicolosi A, *et al*. Drug use in children: cohort study in three European countries. *BMJ* 2008; **337**: a2245.

22 Du Y, Knopf H. Self-medication among children and adolescents in Germany: results of the National Health Survey for Children and Adolescents (KiGGS). *Br J Clin Pharmacol* 2009; **68**(4): 599–608.

23 Yin HS, Parker RM, Sanders LM, *et al*. Effect of medication label units of measure on parent choice of dosing tool: a randomized experiment. *Acad Pediatr* 2016; **16**(8): 734–41.

24 Yin HS, Parker RM, Sanders LM, *et al*. Pictograms, units and dosing tools, and parent medication errors: a randomized study. *Pediatrics* 2017; **140**(1): pii: e20163237.

25 Lammer EJ, Chen DT, Hoar RM, *et al*. Retinoic acid embryopathy. *N Engl J Med* 1985; **313**(14): 837–41.

26 Centers for Disease Control. Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. *Morb Mort Wkly Rep* 1992; **41**: 1–7.

27 Hernandez-Diaz S, Werler MM, Walker AM, Mitchell AA. Neural tube defects in relation to use of folic acid antagonists during pregnancy. *Am J Epidemiol* 2001; **153**(10): 961–8.

28 Scialli AR. Paroxetine exposure during pregnancy and cardiac malformations. *Birth Defects Res A Clin Mol Teratol* 2010; **88**(3): 175–7.

29 Centers for Disease Control. Childhood obesity facts. 2018. https://www.cdc.gov/healthyschools/obesity/facts.htm (accessed April 2018).

30 World Health Organization. Nutrition Landscape Information System (NLIS) country profile indicators interpretation guide. 2010. http://www.who.int/nutrition/nlis_interpretation_guide.pdf (accessed March 2018).

31 Marcovecchio ML, Mohn A, Chiarelli F. Inflammatory cytokines and growth in childhood. *Curr Opin Endocrinol Diabetes Obes* 2012; **19**(1): 57–62.

32 Wells JC. Worldwide variability in growth and its association with health: incorporating body composition, developmental plasticity, and intergenerational effects. *Am J Hum Biol* 2017; **29**(2).

33 Largent EA. Recently proposed changes to legal and ethical guidelines governing human subjects research. *J Law Biosci* 2016; **3**(1): 206–16.

34 Gupta A, Mathad JS, Abdel-Rahman SM, *et al*. Toward earlier inclusion of pregnant and postpartum women in tuberculosis drug trials: consensus statements from an international expert panel. *Clin Infect Dis* 2016; **62**(6): 761–9.

35 National Academies of Sciences Engineering and Medicine. *Integrating Clinical Research into Epidemic Response: The Ebola Experience*. Washington, DC: National Academies Press, 2017.

36 Bracken-Roche D, Bell E, Macdonald ME, Racine E. The concept of "vulnerability" in research ethics: an in-depth analysis of policies and guidelines. *Health Res Policy Syst* 2017; **15**(1): 8.

37 van der Zande ISE, van der Graaf R, Oudijk MA, van Delden JJM. Vulnerability of pregnant women in clinical research. *J Med Ethics* 2017; **43**(10): 657–63.

38 Krubiner CB, Faden RR. Pregnant women should not be categorised as a "vulnerable population" in biomedical research studies: ending a vicious cycle of "vulnerability." *J Med Ethics* 2017; **43**(10): 664–5.

39 Sheffield JS, Siegel D, Mirochnick M, *et al*. Designing drug trials: considerations for pregnant women. *Clin Infect Dis* 2014; **59**(Suppl 7): S437–S444.

40 Klein DB, Tabarrok A. History of federal regulation: Kefauver-Harris Amendments of 1962. 2018. http://www.fdareview.org/01_history.php#p10 (accessed June 2018).

41 Merkatz RB, Junod SW. Historical background of changes in FDA policy on the study and evaluation of drugs in women. *Acad Med* 1994; **69**(9): 703–7.

42 US Food and Drug Administration. Guidance for industry: establishing pregnancy exposure registries. Rockville, MD: US Department of Health and Human Services, 2002.

43 European Parliament and the Council of the European Union. Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. No 536/2014. 2014.

44 Saint-Raymond A, de Vries CS. Medicine safety in pregnancy and ambitions for the EU medicine regulatory framework. *Clin Pharmacol Ther* 2016; **100**(1): 21–3.

45 US Food and Drug Administration, US Department of Health and Human Services. Content and format of labeling for human prescription drug and biological products; requirements for pregnancy and lactation labeling. Final rule. *Fed Regist* 2014; **79**(233): 72063–103.

46 Eunice Kennedy Shriver National Institute of Child Health and Human Development. Task Force on Research Specific to Pregnant Women and Lactating Women (PRGLAC). 2018. https://www.nichd.nih.gov/About/Advisory/PRGLAC (accessed April 2018).

47 Council for International Organizations of Medical Sciences. International ethical guidelines for health-related research involving humans. Geneva: Council for International Organizations of Medical Sciences, 2016. https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf (accessed May 2018).

48 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Welcome to ICH official website. 2018. http://www.ich.org/home.html (accessed September 2018).

49 Knutsen A, Butler AS, Vanchieri C. *Addressing the Barriers to Pediatric Drug Development: Workshop Summary*. Washington, DC: National Academies Press, 2008.

50 European Parliament and the Council of the European Union. Regulation (EC) No 1901/2006 on medicinal products for paediatric use. 2006.

51 Benjamin DK, Jr., Smith PB, Murphy MD, *et al*. Peer-reviewed publication of clinical trials completed for pediatric exclusivity. *JAMA* 2006; **296**(10): 1266–73.

52 Bourgeois FT, Hwang TJ. The Pediatric Research Equity Act moves into adolescence. *JAMA* 2017; **317**(3): 259–60.

53 Bourgeois FT, Hwang TJ. Improving the study of new medicines for children with rare diseases. *JAMA Pediatr* 2018; **172**(1): 7–9.

54 Wilson JG. Evaluation of human teratologic risk in animals. In Lee DH, Hewson EW, Okun D, eds. *Environment and Birth Defects*. New York: Academic Press, 1973, pp. 146–60.

55 Finer L, Henshaw S. Disparities in rates of unintended pregnancy in the United States, 1994 and 2001. *Perspect Sex Reprod Health.* 2006; **38**: 90–6.

56 Patorno E, Huybrechts KF, Bateman BT, *et al*. Lithium use in pregnancy and the risk of cardiac malformations. *N Engl J Med* 2017; **376**(23): 2245–54.

57 Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016; **79**: 70–5.

58 Liang K-Y, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1): 13–22.

59 US Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Pediatric information incorporated into human prescription drug and biological products labeling. 2013. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM341394.pdf (accessed June 2017).

60 National Institutes of Health. Inclusion of children in clinical research: change in NIH definition [NOT-OD-16-010]. 2015. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-010.html (accessed June 2017).

61 Johnson SB, Blum RW, Giedd JN. Adolescent maturity and the brain: the promise and pitfalls of neuroscience research in adolescent health policy. *J Adolesc Health* 2009; **45**(3): 216–21.

62 Hardin AP, Hackell JM, Committee on Practice and Ambulatory Medicine. Age limit of pediatrics. *Pediatrics* 2017; **140**(3): pii: e20172151.

63 US Census Bureau. 2010 census definitions of subject characteristics. 2010. https://www.census.gov/prod/cen2010/doc/sf1.pdf#page=210 (accessed April 2018).

64 International Conference on Harmonisation. Clinical investigation of medicinal products in the pediatric population. 2000. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E11/Step4/E11_Guideline.pdf (accessed April 2018).

65 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP): Product- or population-specific considerations IV: paediatric population. 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/08/WC500232769.pdf (accessed April 2018).

66 Contopoulos-Ioannidis DG, Seto I, Hamm MP, *et al*. Empirical evaluation of age groups and age-subgroup analyses in pediatric randomized trials and pediatric meta-analyses. *Pediatrics* 2012; **129**(Suppl 3): S161–S184.

67 Palmsten K, Hernández-Díaz S, Chambers CD, *et al*. The most commonly dispensed prescription medications among pregnant women enrolled in the United States Medicaid program. *Obstetr Gynecol* 2015; **126**(3): 465–73.

68 Dolk H, Loane M, Garne E. The prevalence of congenital anomalies in Europe. *Adv Exp Med Biol* 2010; **686**: 349–64.

69 Agnello V. Association of systemic lupus erythematosus and SLE-like syndromes with hereditary and acquired complement deficiency states. *Arthritis Rheum* 1978; **21**(5 Suppl): S146–S152.

70 Parker SE, Mai CT, Canfield MA, *et al*. Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006. *Birth Def Res A Clin Mol Teratol* 2010; **88**(12): 1008–16.

71 Yoon PW, Bresee JS, Olney RS, James LM, Khoury MJ. Epidemiology of biliary atresia: a population-based study. *Pediatrics* 1997; **99**(3): 376–82.

72 Coloma PM, Schuemie MJ, Trifiro G, *et al*. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011; **20**(1): 1–11.

73 Sevelsted A, Stokholm J, Bønnelykke K, Bisgaard H. Cesarean section and chronic immune disorders. *Pediatrics* 2015; **135**(1): e92–e98.

74 Yang MS, Kim JY, Kim BK, *et al*. True rise in anaphylaxis incidence: epidemiologic study based on a national health insurance database. *Medicine (Baltimore)* 2017; **96**(5): e5750.

75 Mitchell AA. Systematic identification of drugs that cause birth defects: a new opportunity. *N Engl J Med* 2003; **349**(26): 2556–9.

76 Valkhoff VE, Schade R, t Jong GW, *et al*. Population-based analysis of non-steroidal anti-inflammatory drug use among children in four European countries in the SOS project: what size of data platforms and which study designs do we need to assess safety issues? *BMC Pediatr* 2013; **13**: 192.

77 de Bie S, Coloma PM, Ferrajolo C, *et al*. The role of electronic healthcare record databases in paediatric drug safety surveillance: a retrospective cohort study. *Br J Clin Pharmacol* 2015; **80**(2): 304–14.

78 Rothman KJ. Six persistent research misconceptions. *J Gen Intern Med* 2014; **29**(7): 1060–4.

79 Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; **1**(1): 43–6.

80 Vernacchio L, Kelly JP, Kaufman DW, Mitchell AA. Medication use among children <12 years of age in the United States: results from the Slone Survey. *Pediatrics* 2009; **124**(2): 446–54.

81 Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989; **129**(2): 415–21.

82 Weiss NS. Should we consider a subject's knowledge of the etiologic hypothesis in the analysis of case-control studies? *Am J Epidemiol* 1994; **139**(3): 247–9.

83 Drews CD, Kraus JF, Greenland S. Recall bias in a case–control study of sudden infant death syndrome. *Int J Epidemiol* 1990; **19**(2): 405–11.

84 Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**(4): 670–6.

85 Mitchell AA, Rosenberg L, Shapiro S, Slone D. Birth defects related to bendectin use in pregnancy. I. Oral clefts and cardiac defects. *JAMA* 1981; **245**(22): 2311–14.

86 Garg M, Garrison L, Leeman L, *et al*. Validity of self-reported drug use information among pregnant women. *Matern Child Health J* 2016; **20**(1): 41–7.

87 Delaney-Black V, Chiodo LM, Hannigan JH, *et al*. Just say "I don't": lack of concordance between teen report and biological measures of drug use. *Pediatrics* 2010; **126**(5): 887–93.

88 Suissa S, Dell'aniello S, Vahey S, Renoux C. Time-window bias in case–control studies: statins and lung cancer. *Epidemiology* 2011; **22**(2): 228–31.

89 Hernández-Díaz S. Name of the bias and sex of the angels. *Epidemiology* 2011; **22**(2): 232–3.

90 Yudkin PL, Wood L, Redman CW. Risk of unexplained stillbirth at different gestational ages. *Lancet* 1987; **1**(8543): 1192–4.

91 Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. *Am J Epidemiol* 2011; **174**(9): 1062–8.

92 Platt RW. The fetuses-at-risk-approach: an evolving paradigm. In Buck Louis G, Platt R, eds. *Reproductive and Perinatal Epidemiology*. New York: Oxford University Press, 2011, pp. 262–75.

93 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009; **170**(2): 244–56.

94 Edwards JK, Hester LL, Gokhale M, Lesko CR. Methodologic issues when estimating risks in

pharmacoepidemiology. *Curr Epidemiol Rep* 2016; **3**(4): 285–96.

95 Hernandez-Diaz S, Werler MM, Walker AM, Mitchell AA. Folic acid antagonists during pregnancy and the risk of birth defects. *N Engl J Med* 2000; **343**(22): 1608–14.

96 Horton DB, Haynes K, Denburg MR, *et al*. Oral glucocorticoid use and osteonecrosis in children and adults with chronic inflammatory diseases: a population-based cohort study. *BMJ Open* 2017; **7**(7): e016788.

97 Tambralli A, Beukelman T, Weiser P, Atkinson TP, Cron RQ, Stoll ML. High doses of infliximab in the management of juvenile idiopathic arthritis. *J Rheumatol* 2013; **40**(10): 1749–55.

98 Shaw K, Amstutz U, Hildebrand C, *et al*. VKORC1 and CYP2C9 genotypes are predictors of warfarin-related outcomes in children. *Pediatr Blood Cancer* 2014; **61**(6): 1055–62.

99 Cooper W, Hernandez-Diaz S, Arbogast P, *et al*. Major congenital malformations after first-trimester exposure to ACE inhibitors. *N Engl J Med* 2006; **354**: 2443–51.

100 Cooper W, Hernandez-Diaz S, Ray W. Positive predictive value of computerized databases for birth defects in a medicaid population. *Pharmacoepidemiol Drug Saf* 2004; **13**: S1–S334.

101 Schelleman H, Bilker WB, Strom BL, *et al*. Cardiovascular events and death in children exposed and unexposed to ADHD agents. *Pediatrics* 2011; **127**(6): 1102–10.

102 Cooper WO, Hernandez-Diaz S, Gideon P, *et al*. Positive predictive value of computerized records for major congenital malformations. *Pharmacoepidemiol Drug Saf* 2008; **17**(5): 455–60.

103 Robins J. Causal inference from complex longitudinal data. In Berkane M, ed., *Lecture Notes in Statistics*. New York: Springer, 1997, pp. 67–117.

104 Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**(5): 550–60.

105 Khoury MJ, Moore CA, James LM, Cordero JF. The interaction between dysmorphology and epidemiology: methodologic issues of lumping and splitting. *Teratology* 1992; **45**: 133–8.

106 Rasmussen SA, Olney RS, Holmes LB, Lin AE, Keppler-Noreuil KM, Moore CA. Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Def Res A Clin Mol Teratol* 2003; **67**(3): 193–201.

107 Kirby ML, Gale TF, Stewart DE. Neural crest cells contribute to normal aorticopulmonary septation. *Science* 1983; **220**: 1059–61.

108 Werler MM, Sheehan J, Mitchell A. Association of vasoconstrictive exposures with risks of gastroschisis and small intestinal atresia. *Epidemiology* 2003; **13**(3): 349–54.

109 Ferencz C, Correa-Villasenor A, Loffredo C, Wilson PD, eds. *Genetic and Environmental Risk Factors of Major Cardiovascular Malformations: The Baltimore–Washington Infant Study: 1981–1989*. Armonk, NY: Futura, 1997.

110 Holmes LB, Westgate MN. Inclusion and exclusion criteria for malformations in newborn infants exposed to potential teratogens. *Birth Def Res A Clin Mol Teratol* 2011; **91**(9): 807–12.

111 Holmes LB. Need for inclusion and exclusion criteria for the structural abnormalities recorded in children born from exposed pregnancies. *Teratology* 1999; **59**: 1–2.

112 Cragan JD, Gilboa SM. Including prenatal diagnoses in birth defects monitoring: experience of the Metropolitan Atlanta Congenital Defects Program. *Birth Def Res A Clin Mol Teratol* 2009; **85** (1): 20–9.

113 Correa A, Cragan JD, Kucik JE, *et al*. Reporting birth defects surveillance data 1968–2003. *Birth Def Res A Clin Mol Teratol* 2007; **79**(2): 65–186.

114 Honein M, Paulozzi LJ, Cragan JD, Correa A. Evaluation of selected characteristics of pregnancy drug registries. *Teratology* 1999; **60**: 356–64.

115 MacDorman MF, Gregory EC. Fetal and perinatal mortality: United States, 2013. *Natl Vital Stat Rep* 2015; **64**(8): 1–24.

116 Margulis AV, Mittleman MA, Glynn RJ, Holmes LB, Hernandez-Diaz S. Effects of gestational age at enrollment in pregnancy exposure registries. *Pharmacoepidemiol Drug Saf* 2015; **24**(4): 343–52.

117 Strom BL, Carson JL, Schinnar R, Shaw M. Is cimetidine associated with neutropenia? *Am J Med* 1995; **99**(3): 282–90.

118 Thomas S, Edwards C, Smeeth L, Cooper C, Hall A. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? *Arthritis Rheum* 2008; **59**(9): 1314–21.

119 Salerno M, Micillo M, Di Maio S, *et al*. Longitudinal growth, sexual maturation and final height in patients with congenital hypothyroidism detected by neonatal screening. *Eur J Endocrinol* 2001; **145**(4): 377–83.

120 Walters TD, Kim MO, Denson LA, *et al*. Increased effectiveness of early therapy with anti-tumor necrosis factor-alpha vs an immunomodulator in children with Crohn's disease. *Gastroenterology* 2014; **146**(2): 383–91.

121 Kearsley-Fleet L, Hyrich KL, Davies R, *et al*. Growth in children and adolescents with juvenile idiopathic arthritis over 2 years of treatment with etanercept: results from the British Society for Paediatric and Adolescent Rheumatology Etanercept Cohort Study. *Rheumatology (Oxford)* 2015; **54**(7): 1279–85.

122 De Leonibus C, Attanasi M, Roze Z, *et al*. Influence of inhaled corticosteroids on pubertal growth and final height in asthmatic children. *Pediatr Allergy Immunol* 2016; **27**(5): 499–506.

123 Poulton AS, Bui Q, Melzer E, Evans R. Stimulant medication effects on growth and bone age in children with attention-deficit/hyperactivity disorder: a prospective cohort study. *Int Clin Psychopharmacol* 2016; **31**(2): 93–9.

124 Scott FI, Horton DB, Mamtani R, *et al*. Administration of antibiotics to children before age 2 years increases risk for childhood obesity. *Gastroenterology* 2016; **151**(1): 120–9, e125.

125 Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology (Cambridge, Mass)* 2009; **20**(4): 512–22.

126 Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Philadelphia, PA: Lippincott-Raven, 1998.

127 Walker AM. *Observation and Inference: An Introduction to the Methods of Epidemiology*. Newton Lower Falls, MA: Epidemiology Resources, 1991.

128 Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014; **43**(6): 1969–85.

129 Huybrechts KF, Palmsten K, Avorn J, *et al*. Antidepressant use in pregnancy and the risk of cardiac defects. *N Engl J Med* 2014; **370**(25): 2397–407.

130 Koren G. The effect of ascertainment bias in evaluating gestational antidepressant exposure. *J Popul Ther Clin Pharmacol* 2011; **18**: e174–e175.

131 Dufendach KR, Eichenberger JA, McPheeters ML, *et al*. *Core Functionality in Pediatric Electronic Health Records*.

Rockville, MD: Agency for Healthcare Research and Quality, 2015.

132 Kinlaw AC, Sturmer T, Lund JL, *et al*. Trends in antibiotic use by birth season and birth year. *Pediatrics* 2017; **140**(3): e20174250.

133 Mattano LA, Sather HN, Trigg ME, Nachman JB. Osteonecrosis as a complication of treating acute lymphoblastic leukemia in children: a report from the Children's Cancer Group. *J Clin Oncol* 2000; **18**(18): 3262–72.

134 Tsampalieros A, Lam CKL, Spencer JC, *et al*. Long-term inflammation and glucocorticoid therapy impair skeletal modeling during growth in childhood Crohn disease. *J Clin Endocrinol Metab* 2013; **98**(8): 3438–45.

135 Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004; **15**(5): 615–25.

136 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**: 915–20.

137 Park Y, Hernandez-Diaz S, Bateman B, *et al*. Continuation of atypical antipsychotic medication during early pregnancy and the risk of gestational diabetes. *Am J Psychiatry* 2018: **175**(6): 564–74.

138 Kinlaw AC, Buckley JP, Engel SM, Poole C, Brookhart MA, Keil AP. Left truncation bias to explain the protective effect of smoking on preeclampsia: potential, but how plausible? *Epidemiology* 2017; **28**(3): 428–34.

139 Svensson E, Ehrenstein V, Norgaard M, *et al*. Estimating the proportion of all observed birth defects occurring in pregnancies terminated by a second-trimester abortion. *Epidemiology* 2014; **25**(6): 866–71.

140 Friedberg MK, Silverman NH, Moon-Grady AJ, *et al*. Prenatal detection of congenital heart disease. *J Pediatr* 2009; **155**(1): 26–31, e21.

141 Kieler H, Malm H, Artama M, *et al*. Use of antidepressants and association with elective termination of pregnancy: population based case-control study. *BJOG* 2015; **122**(12): 1618–24.

142 Lash TL, Schmidt M, Jensen AO, Engebjerg MC. Methods to apply probabilistic bias analysis to summary estimates of association. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 638–44.

143 Greenland S. Multiple-bias modelling for analysis of observational data. *J R Statist Soc A* 2005; **168**: 267–306.

144 Mills JL, Rhoads GG, Simpson JL, *et al*. The absence of a relation between the periconceptional use of vitamins and neural-tube defects. *N Engl J Med* 1989; **321**: 430–5.

145 Bracken MB. Methodologic issues in the epidemiologic investigation of drug-induced congenital malformations. In Bracken MB ed., *Perinatal Epidemiology*. New York: Oxford University Press, 1984, pp. 423–49.

146 Lieff S, Olsham A, Werler M, Savitz DA, Mitchell AA. Selection bias and the use of controls with malformations in case-control studies of birth defects. *Epidemiology* 1999; **10**: 238–41.

147 Lin KJ, Mitchell AA, Yau WP, Louik C, Hernandez-Diaz S. Maternal exposure to amoxicillin and the risk of oral clefts. *Epidemiology* 2012; **23**(5): 699–705.

148 Greenland S. The effect of misclassification in matched-pair case-control studies. *Am J Epidemiol* 1982; **116**: 402–6.

149 Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002; **155**(2): 176–84.

150 Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight "paradox" uncovered? *Am J Epidemiol* 2006; **164**(11): 1115–20.

151 Woolfenden S, Eapen V, Axelsson E, *et al*. Who is our cohort: recruitment, representativeness, baseline risk and

retention in the "Watch Me Grow" study? *BMC Pediatr* 2016; **16**: 46.

152 Johnson SB, Lynch KF, Baxter J, *et al*. Predicting later study withdrawal in participants active in a longitudinal birth cohort study for 1 year: The TEDDY study. *J Pediatr Psychol* 2016; **41**(3): 373–83.

153 Altit G, Basso O, Grandi SM, Yang S. Letter to the editor. *J Perinatol* 2018; **38**(6): 767.

154 Nelin TD, Pena E, Giacomazzi T, *et al*. Outcomes following indomethacin prophylaxis in extremely preterm infants in an all-referral NICU. *J Perinatol* 2017; **37**(8): 932–7.

155 dosReis S, Zito JM, Safer DJ, Soeken KL. Mental health services for youths in foster care and disabled youths. *Am J Public Health* 2001; **91**(7): 1094–9.

156 Szilagyi MA, Rosen DS, Rubin D, *et al*. Health care issues for children and adolescents in foster care and kinship care. *Pediatrics* 2015; **136**(4): e1142–e1166.

157 Rubin DM, Pati S, Luan X, Alessandrini EA. A sampling bias in identifying children in foster care using Medicaid data. *Ambul Pediatr* 2005; **5**(3): 185–90.

158 Raghavan R, Brown DS, Allaire BT. Can Medicaid claims validly ascertain foster care status? *Child Maltreat* 2017; **22**(3): 227–35.

159 Olsen J, Melbye M, Olsen SF, *et al*. The Danish National Birth Cohort: its background, structure and aim. *Scand J Public Health* 2001; **29**(4): 300–7.

160 Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* 2006; **35**(5): 1146–50.

161 Steinhorn R, McPherson C, Anderson PJ, Neil J, Doyle LW, Inder T. Neonatal morphine exposure in very preterm infants-cerebral development and outcomes. *J Pediatr* 2015; **166**(5): 1200–7, e1204.

162 Batton B, Li L, Newman NS, *et al*. Early blood pressure, antihypotensive therapy and outcomes at 18–22 months' corrected age in extremely preterm infants. *Arch Dis Child Fetal Neonatal Ed* 2016; **101**(3): F201–F206.

163 Uusitalo U, Liu X, Yang J, *et al*. Association of early exposure of probiotics and islet autoimmunity in the TEDDY study. *JAMA Pediatr* 2016; **170**(1): 20–8.

164 Wong CJ, Moxey-Mims M, Jerry-Fluker J, Warady BA, Furth SL. CKiD (CKD in children) prospective cohort study: a review of current findings. *Am J Kidney Dis* 2012; **60**(6): 1002–11.

165 Sidhu N, Joffe AR, Doughty P, *et al*. Sepsis after cardiac surgery early in infancy and adverse 4.5-year neurocognitive outcomes. *J Am Heart Assoc* 2015; **4**(8): e001954.

166 Kugathasan S, Denson LA, Walters TD, *et al*. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* 2017; **389**(10080): 1710–18.

167 Guzman J, Oen K, Tucker LB, *et al*. The outcomes of juvenile idiopathic arthritis in children managed with contemporary treatments: results from the ReACCh-Out cohort. *Ann Rheum Dis* 2015; **74**(10): 1854–60.

168 Vernal DL, Kapoor S, Al-Jadiri A, *et al*. Outcome of youth with early-phase schizophrenia-spectrum disorders and psychosis not otherwise specified treated with second-generation antipsychotics: 12 week results from a prospective, naturalistic cohort study. *J Child Adolesc Psychopharmacol* 2015; **25**(7): 535–47.

169 Goldenberg NA, Branchford B, Wang M, Ray C, Jr., Durham JD, Manco-Johnson MJ. Percutaneous mechanical and pharmacomechanical thrombolysis for occlusive deep vein thrombosis of the proximal limb in adolescent subjects: findings from an institution-based prospective inception cohort study of pediatric venous thromboembolism. *J Vasc Interv Radiol* 2011; **22**(2): 121–32.

170 Honein MA, Paulozzi LJ, Cragan JD, Correa A. Evaluation of selected characteristics of pregnancy drug registries. *Teratology* 1999; **60**: 356–64.

171 Kennedy DL, Uhl K, Kweder SL. Pregnancy exposure registries. *Drug Saf* 2004; **27**: 215–28.

172 Mitchell AA. Systematic identification of drugs that cause birth defects: a new opportunity. *New Engl J Med* 2003; **349**: 2556–9.

173 Hernandez-Diaz S, Smith CR, Shen A, *et al*. Comparative safety of antiepileptic drugs during pregnancy. *Neurology* 2012; **78**(21): 1692–9.

174 Natter MD, Quan J, Ortiz DM, *et al*. An i2b2-based, generalizable, open source, self-scaling chronic disease registry. *J Am Med Inform Assoc* 2013; **20**(1): 172–9.

175 Sunseri W, Hyams JS, Lerer T, *et al*. Retrospective cohort study of methotrexate use in the treatment of pediatric Crohn's disease. *Inflamm Bowel Dis* 2014; **20**(8): 1341–5.

176 Oliva-Hemker M, Hutfless S, Al Kazzi ES, *et al*. Clinical presentation and five-year therapeutic management of very early-onset inflammatory bowel disease in a large North American cohort. *J Pediatr* 2015; **167**(3): 527–32, e521–e523.

177 Trautmann A, Bodria M, Ozaltin F, *et al*. Spectrum of steroid-resistant and congenital nephrotic syndrome in children: the PodoNet registry cohort. *Clin J Am Soc Nephrol* 2015; **10**(4): 592–600.

178 Thurber BW, Carmody D, Tadie EC, *et al*. Age at the time of sulfonylurea initiation influences treatment outcomes in KCNJ11-related neonatal diabetes. *Diabetologia* 2015; **58**(7): 1430–5.

179 Kato S, Yabe H, Takakura H, *et al*. Hematopoietic stem cell transplantation for inborn errors of metabolism: A report from the Research Committee on Transplantation for Inborn Errors of Metabolism of the Japanese Ministry of Health, Labour and Welfare and the Working Group of the Japan Society for Hematopoietic Cell Transplantation. *Pediatr Transplant* 2016; **20**(2): 203–14.

180 Marquant F, Goujon S, Faure L, *et al*. Risk of childhood cancer and socio-economic disparities: results of the French nationwide study Geocap 2002–2010. *Paediatr Perinat Epidemiol* 2016; **30**(6): 612–22.

181 Haessler F, Gaese F, Huss M, *et al*. Characterization, treatment patterns, and patient-related outcomes of patients with Fragile X syndrome in Germany: final results of the observational EXPLAIN-FXS study. *BMC Psychiatr* 2016; **16**(1): 318.

182 Seidlin M, Holzman R, Knight P, *et al*. Characterization and utilization of an international neurofibromatosis web-based, patient-entered registry: an observational study. *PLoS One* 2017; **12**(6): e0178639.

183 Hernandez-Diaz S. Prescription of medications during pregnancy: accidents, compromises, and uncertainties. *Pharmacoepidemiol Drug Saf* 2006; **15**(9): 613–17.

184 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**(4): 323–37.

185 Charlton RA, Cunnington M, de Vries CS, Weil JG. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf* 2008; **31**: 39–51.

186 Palmsten K, Huybrechts KF, Mogun H, *et al*. Harnessing the Medicaid Analytic eXtract (MAX) to evaluate medications in pregnancy: design considerations. *PLoS One* 2013; **8**(6): e67405.

187 Cea-Soriano L, Garcia Rodriguez LA, Fernandez Cantero O, Hernandez-Diaz S.

Challenges of using primary care electronic medical records in the UK to study medications in pregnancy. *Pharmacoepidemiol Drug Saf* 2013; **22**(9): 977–85.

188 Harron K, Gilbert R, Cromwell D, van der Meulen J. Linking data for mothers and babies in de-identified electronic health data. *PLoS One* 2016; **11**(10): e0164667.

189 Andrade SE, Davis RL, Cheetham TC, *et al*. Medication exposure in pregnancy risk evaluation program. *Mat Child Health J* 2012; **16**(7): 1349–54.

190 Li Q, Andrade SE, Cooper WO, *et al*. Validation of an algorithm to estimate gestational age in electronic health plan databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(5): 524–32.

191 Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(1): 16–24.

192 Andrade SE, Toh S, Houstoun M, *et al*. Surveillance of medication use during pregnancy in the Mini-Sentinel program. *Mat Child Health J* 2016; **20**(4): 895–903.

193 Huybrechts K, Bröms G, Brix Christensen L, *et al*. Association between methylphenidate and amphetamine use in pregnancy and risk of congenital malformations: a cohort study from the International Pregnancy Safety Study Consortium. *JAMA Psychiatr* 2018; **75**(2): 167–75.

194 Sinclair S, Cunnington M, Messenheimer J, *et al*. Advantages and problems with pregnancy registries: observations and surprises throughout the life of the International Lamotrigine Pregnancy Registry. *Pharmacoepidemiol Drug Saf* 2014; **23**(8): 779–86.

195 Kenyon CC, Rubin DM, Zorc JJ, Mohamad Z, Faerber JA, Feudtner C. Childhood asthma hospital discharge medication fills and risk of subsequent readmission. *J Pediatr* 2015; **166**(5): 1121–7.

196 Furu K, Karlstad O, Zoega H, *et al*. Utilization of stimulants and atomoxetine for attention-deficit/hyperactivity disorder among 5.4 million children using population-based longitudinal data. *Basic Clin Pharmacol Toxicol* 2017; **120**(4): 373–9.

197 Chen VC, Yang YH, Liao YT, *et al*. The association between methylphenidate treatment and the risk for fracture among young ADHD patients: a nationwide population-based study in Taiwan. *PLoS One* 2017; **12**(3): e0173762.

198 Beukelman T, Xie F, Chen L, *et al*. Risk of malignancy associated with paediatric use of tumour necrosis factor inhibitors. *Ann Rheum Dis* 2018; **77**(7): 1012–16.

199 Rani FA, Byrne P, Cranswick N, Murray ML, Wong IC. Mortality in children and adolescents prescribed antipsychotic medication: a retrospective cohort study using the UK general practice research database. *Drug Saf* 2011; **34**(9): 773–81.

200 Eberly MD, Eide MB, Thompson JL, Nylund CM. Azithromycin in early infancy and pyloric stenosis. *Pediatrics* 2015; **135**(3): 483–8.

201 Horton DB, Scott FI, Haynes K, *et al*. Antibiotic exposure, infection, and the development of pediatric psoriasis: a nested case-control study. *JAMA Dermatol* 2016; **152**(2): 1–9.

202 Ortqvist AK, Lundholm C, Kieler H, *et al*. Antibiotics in fetal and early life and subsequent childhood asthma: nationwide population based study with sibling analysis. *BMJ* 2014; **349**: g6979.

203 Horton DB, Scott FI, Haynes K, *et al*. Antibiotic exposure and juvenile idiopathic arthritis: a case-control study. *Pediatrics* 2015; **136**(2): e333–e343.

204 Ceci A, Giaquinto C, Aboulker JP, *et al*. The Task-force in Europe for Drug Development for the Young (TEDDY) network of

excellence. *Paediatr Drugs* 2009; **11**(1): 18–21.

205 Swerdlow AJ, Cooke R, Beckers D, *et al.* Cancer risks in patients treated with growth hormone in childhood: the SAGhE European cohort study. *J Clin Endocrinol Metab* 2017; **102**(5): 1661–72.

206 Fiks AG, Grundmeier RW, Steffes J, *et al.* Comparative effectiveness research through a collaborative electronic reporting consortium. *Pediatrics* 2015; **136**(1): e215–e224.

207 Czaja AS, Ross ME, Liu W, *et al.* Electronic health record (EHR) based postmarketing surveillance of adverse events associated with pediatric off-label medication use: a case study of short-acting beta-2 agonists and arrhythmias. *Pharmacoepidemiol Drug Saf* 2018; **27**(7): 815–22.

208 Forrest CB, Margolis PA, Bailey LC, *et al.* PEDSnet: a national pediatric learning health system. *J Am Med Inform Assoc* 2014; **21**(4): 602–6.

209 Kavcic M, Fisher BT, Seif AE, *et al.* Leveraging administrative data to monitor rituximab use in 2875 patients at 42 freestanding children's hospitals across the United States. *J Pediatr* 2013; **162**(6): 1252–8, e1251.

210 McQueen M, Rojas J, Sun SC, *et al.* Safety and long term outcomes with high flow nasal cannula therapy in neonatology: a large retrospective cohort study. *J Pulm Respir Med* 2014; **4**(6): 216.

211 Dai D, Feinstein JA, Morrison W, Zuppa AF, Feudtner C. Epidemiology of polypharmacy and potential drug-drug interactions among pediatric patients in ICUs of U.S. children's hospitals. *Pediatr Crit Care Med* 2016; **17**(5): e218–e228.

212 Ericson JE, Gostelow M, Autmizguine J, *et al.* Safety of high-dose acyclovir in infants with suspected and confirmed neonatal herpes simplex virus infections. *Pediatr Infect Dis J* 2017; **36**(4): 369–73.

213 James KE, Xiao R, Merkel PA, Weiss PF. Variation in the treatment of children hospitalized with antineutrophil cytoplasmic antibody-associated vasculitis in the US. *Arthritis Care Res (Hoboken)* 2017; **69**(9): 1377–83.

214 Lindblad F, Ringback Weitoft G, Hjern A. Maternal and paternal psychopathology increases risk of offspring ADHD equally. *Epidemiol Psychiatr Sci* 2011; **20**(4): 367–72.

215 Palmer N, Beam A, Agniel D, *et al.* Association of sex with recurrence of autism spectrum disorder among siblings. *JAMA Pediatr* 2017; **171**(11): 1107–12.

216 Winterstein AG, Choi Y, Meissner HC. Association of age with risk of hospitalization for respiratory syncytial virus in preterm infants with chronic lung disease. *JAMA Pediatr* 2018; **172**(2): 154–60.

217 Brikell I, Ghirardi L, D'Onofrio BM, *et al.* Familial liability to epilepsy and attention-deficit/hyperactivity disorder: a nationwide cohort study. *Biol Psychiatr* 2018; **83**(2): 173–80.

218 van der Schans J, Vardar S, Cicek R, *et al.* An explorative study of school performance and antipsychotic medication. *BMC Psychiatr* 2016; **16**(1): 332.

219 van der Schans J, Cicek R, Vardar S, *et al.* Methylphenidate use and school performance among primary school children: a descriptive study. *BMC Psychiatr* 2017; **17**(1): 116.

220 Karayurt O, Ordin YS, Unek T, Astarcioglu I. Immunosuppressive medication adherence, therapeutic adherence, school performance, symptom experience, and depression levels in patients having undergone a liver transplant during childhood. *Exp Clin Transplant* 2015; **13**(3): 247–55.

221 Liew Z, Ritz B, Virk J, Arah OA, Olsen J. Prenatal use of acetaminophen and child IQ: a Danish cohort study. *Epidemiology* 2016; **27**(6): 912–18.

222 Bromley RL, Calderbank R, Cheyne CP, *et al*. Cognition in school-age children exposed to levetiracetam, topiramate, or sodium valproate. *Neurology* 2016; **87**(18): 1943–53.

223 Werler MM, Louik C, Mitchell AA. Case–control studies for identifying novel teratogens. *Am J Med Genet C Semin Med Genet* 2011; **157C**(3): 201–8.

224 Bertuola F, Morando C, Menniti-Ippolito F, *et al*. Association between drug and vaccine use and acute immune thrombocytopenia in childhood: a case–control study in Italy. *Drug Saf* 2010; **33**(1): 65–72.

225 Rockenbauer M, Olsen J, Czeizel A, Pedersen L, Sorensen H, Group E. Recall bias in a case–control surveillance system on the use of medicine during pregnancy. *Epidemiology* 2001; **12**: 461–6.

226 Drews CD, Greenland S. The impact of differential recall on the results of case–control studies. *Int J Epidemiol* 1990; **19**(4): 1107–12.

227 Luteijn JM, Morris JK, Garne E, *et al*. EUROmediCAT signal detection: a systematic method for identifying potential teratogenic medication. *Br J Clin Pharmacol* 2016; **82**(4): 1110–22.

228 Gili JA, Poletta FA, Gimenez LG, *et al*. Descriptive analysis of high birth prevalence rate geographical clusters of congenital anomalies in South America. *Birth Def Res A Clin Mol Teratol* 2016; **106**(4): 257–66.

229 Hernandez-Diaz S, Hernan MA, Meyer K, Werler MM, Mitchell AA. Case-crossover and case-time-control designs in birth defects epidemiology. *Am J Epidemiol* 2003; **158**(4): 385–91.

230 Meyer KA, Williams P, Hernandez-Diaz S, Cnattingius S. Smoking and the risk of oral clefts: exploring the impact of study designs. *Epidemiology* 2004; **15**(6): 671–8.

231 Donovan SJ, Susser E. Commentary: advent of sibling designs. *Int J Epidemiol* 2011; **40**(2): 345–9.

232 Shin JY, Roughead EE, Park BJ, Pratt NL. Cardiovascular safety of methylphenidate among children and young people with attention-deficit/hyperactivity disorder (ADHD): nationwide self controlled case series study. *BMJ* 2016; **353**: i2550.

233 Tate JE, Mwenda JM, Armah G, *et al*. Evaluation of Intussusception after monovalent rotavirus vaccination in Africa. *N Engl J Med* 2018; **378**(16): 1521–8.

234 Ueyama H, Hinotsu S, Tanaka S, *et al*. Application of a self-controlled case series study to a database study in children. *Drug Saf* 2014; **37**(4): 259–68.

235 Sun LS, Li G, Miller TL, *et al*. Association between a single general anesthesia exposure before age 36 months and neurocognitive outcomes in later childhood. *JAMA* 2016; **315**(21): 2312–20.

236 Okubo Y, Michihata N, Morisaki N, *et al*. Association between dose of glucocorticoids and coronary artery lesions in Kawasaki disease. *Arthritis Care Res (Hoboken)* 2018; **70**(7):1052–7.

237 Slaughter JL, Reagan PB, Newman TB, Klebanoff MA. Comparative effectiveness of nonsteroidal anti-inflammatory drug treatment vs no treatment for patent ductus arteriosus in preterm infants. *JAMA Pediatr* 2017; **171**(3): e164354.

238 Anglemyer A, Sturt A, Maldonado Y. The effect of combination antiretroviral therapy use among HIV positive children on the hazard of AIDS using calendar year as an instrumental variable. *Curr HIV Res* 2018; **16**(2): 151–7.

239 Ameyaw E, Nguah SB, Ansong D, Page I, Guillerm M, Bates I. The outcome of a test-treat package versus routine outpatient care for Ghanaian children with fever: a pragmatic randomized control trial. *Malar J* 2014; **13**: 461.

240 Chalmers JR, Haines RH, Mitchell EJ, *et al*. Effectiveness and cost-effectiveness of daily all-over-body application of emollient during the first year of life for preventing atopic eczema in high-risk children (the

BEEP trial): protocol for a randomised controlled trial. *Trials* 2017; **18**(1): 343.

241 Franklin JM, Eddings W, Schneeweiss S, Rassen JA. Incorporating linked healthcare claims to improve confounding control in a study of in-hospital medication use. *Drug Saf* 2015; **38**(6): 589–600.

242 Visscher H, Ross CJ, Rassekh SR, *et al*. Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J Clin Oncol* 2012; **30**(13): 1422–8.

243 Ansari AH, Cherian PJ, Dereymaeker A, *et al*. Improved multi-stage neonatal seizure detection using a heuristic classifier and a data-driven post-processor. *Clin Neurophysiol* 2016; **127**(9): 3014–24.

244 Takeuchi M, Inuzuka R, Hayashi T, *et al*. Novel risk assessment tool for immunoglobulin resistance in Kawasaki disease: application using a random forest classifier. *Pediatr Infect Dis J* 2017; **36**(9): 821–6.

245 Brinkman WB, Lipstein EA, Taylor J, *et al*. Design and implementation of a decision aid for juvenile idiopathic arthritis medication choices. *Pediatr Rheumatol Online J* 2017; **15**(1): 48.

23

Study of Biologics and Biosimilars

Jeffrey R. Curtis[1] *and James D. Lewis*[2]

[1] *Division of Clinical Immunology & Rheumatology, University of Alabama at Birmingham, Birmingham, AL, USA*
[2] *Department of Medicine, Division of Gastroenterology, and Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA, USA*

Biologic therapies are complex molecules that include a variety of products, such as recombinant therapeutic proteins, vaccines, blood components, gene therapy, and others. Their components or precursors are derived from living sources. From the perspective of pharmacoepidemiology, biologic drugs are therapeutic agents used to prevent or treat a health condition. They typically take one of three constructs: (i) fusion proteins that link a receptor with a protein (e.g., Fc region of an antibody) or a polyethylene glycol (PEG) fragment, which extends the half-life of the receptor portion of the construct; (ii) custom monoclonal antibodies that can be humanized or chimeric (e.g., with both human and murine components); and (iii) an agent that mimics the human native signaling mechanism, such as erythropoietin, growth hormone, insulin, or human parathyroid hormone.

Biologics are typically more difficult to synthesize than traditional small molecules and require synthesis in a bioreactor. Examples of the molecular constructs for several biologics used for the treatment of autoimmune and inflammatory diseases are shown in Figure 23.1. Currently available biologics are administered parenterally, typically by intravenous infusion or subcutaneous injection; those given by injection are often but not always self-administered by the patient.

In the US, biologics were approved for use in humans as early as 1986 [1], although development efforts accelerated dramatically in the late 1990s. Their clinical indications have expanded from somewhat narrow, niche uses (e.g., organ transplantation, adjunct treatment to percutaneous coronary intervention, treatment of specific types of malignancies) to more widespread use for a variety of autoimmune and inflammatory diseases, including rheumatoid arthritis (RA) and other forms of inflammatory arthritis (e.g., psoriatic arthritis, ankylosing spondylitis, gout), inflammatory bowel disease (IBD), psoriasis, and multiple sclerosis (MS); see Table 23.1. Over time, the indications for biologic use have expanded and now include noninflammatory conditions (e.g., osteoporosis). Although the specific mechanisms of action (MOA), safety concerns, and effects of various biologics are highly variable, the general principles that underlie use of these drugs as related to pharmacoepidemiologic research will be reviewed in this chapter.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

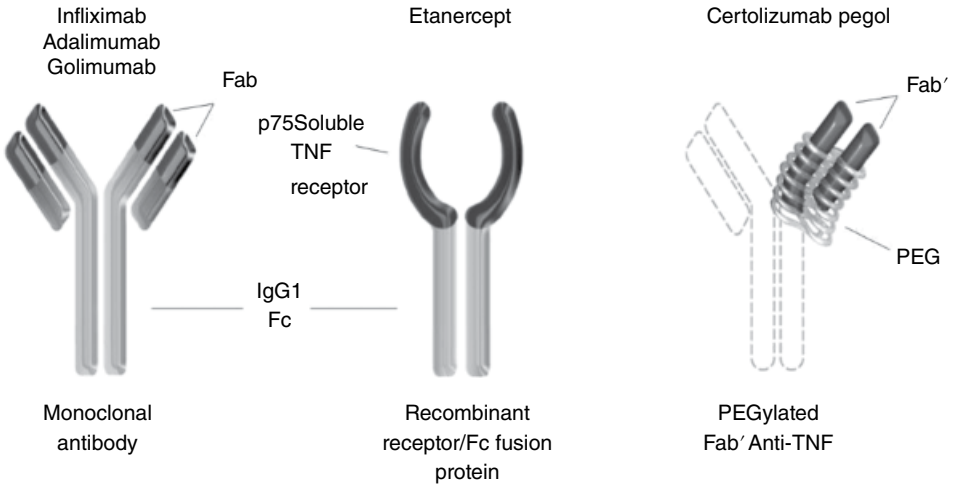


Figure 23.1 Examples of the molecular constructs for several biologics used for the treatment of autoimmune and inflammatory diseases.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

The pharmacoepidemiology questions that surround the use of most biologics are generally similar to those of other immunosuppressant medications. Because many of the disease indications for biologics (e.g., RA, IBD) have a relatively low prevalence (e.g., population prevalence of each of approximately 1%), pharmacoepidemiologic methods to study rare safety events have been particularly valuable to characterize the risks and benefits of these medications. Some safety concerns are common to all biologics, including risks for serious infections (both bacterial infections such as pneumonia and sepsis, as well as rare opportunistic infections such as tuberculosis) and malignancy. Others are more mechanism specific and related to particular effects of biologic pathways (e.g., lipid metabolism) that may increase concern for increased risk for particular types of events (e.g., myocardial infarction). There are a few unique types of adverse reactions that have been associated with biologics that are thought to be related to the consequences of administering large, biologically active proteins. Each of these will be briefly discussed in what follows.

Serious Infections

All medications that suppress the immune system are thought to increase the risk of infection. Unfortunately, it is difficult to quantify how much any specific medication suppresses the immune system. In general, infections that typically only occur in the setting of immunosuppression are referred to as opportunistic infections. Some medications may increase the risk of common infections (e.g., pneumonia), while others predispose to selected types of opportunistic infections. The complex nature of the human immune system results in different patterns of infection based on the aspect of the immune system that is being targeted.

Shortly after the marketing of anti-TNF (tumor necrosis factor) drugs to treat immune-mediated diseases, there was a signal in postmarketing reporting of an increased risk of opportunistic

Table 23.1 Examples of biologics, their mechanism of action, and FDA-approved indications for non-malignant conditions.

MOA	Name (construct)	Examples of FDA-approved indications
Cytokine or cytokine receptor inhibition		
TNF inhibitor (TNFi)	Etanercept (fusion protein)	RA, PsA, AS, JIA, PsO
	Adalimumab (humanized monoclonal antibody)	RA, PsA, AS, JIA, PsO, uveitis
	Infliximab (chimeric monoclonal antibody)	RA, PsA, AS, IBD, PsO
	Golimumab (humanized monoclonal antibody)	RA, PsA, AS, PsO
	Certolizumab (PEGylated protein)	RA, PsA, AS, PsO, IBD
Anti-IL6R	Tocilizumab (monoclonal antibody)	RA, sJIA, GCA
	Sarilumab (monoclonal antibody)	RA
Anti-IL1R	Anakinra (fusion protein)	RA, periodic syndromes
IL-1	Rilonacept (soluble decoy receptor)	Periodic syndromes
Anti-IL-1β	Canakinumb (monoclonal antibody)	sJIA, periodic syndromes, AOSD
IL-17	Secukinumab, Ixekizumab (monoclonal antibody)	PsO, AS, PsA
IL-12/23		PsO
IL-23	Ustekinumab (monoclonal antibody)	PsO, Crohn's disease
	Guselkumab (monoclonal antibody)	PsO
RANK/RANKL	Denosumab	Osteoporosis
Anti-interferon beta-1	Multiple, directed against beta-1a or beta-1b	MS
Other mechanisms of action		
B-cell depletion	Rituximab (monoclonal antibody)	RA, vasculitis
T-cell co-stimulation blockade	Abatacept (fusion protein)	RA, PsA, JIA
B-lymphocyte stimulator antagonist	Belimumab (monoclonal antibody)	SLE
Recombinant mammalian urate oxidase (uricase)	Pegloticase (PEGylated protein)	Gout
Antagonist to α4 integrin	Natalizumab (monoclonal antibody)	MS, IBD
Antagonist to α4β7 integrin	Vedolizumab (monoclonal antibody)	IBD

AOSD, adult-onset Still's disease; AS, ankylosing spondylitis; GCA, giant cell arthritis; IBD, inflammatory bowel disease; JIA, juvenile idiopathic arthritis; MOA, mechanism of action; MS, multiple sclerosis; PsA, psoriatic arthritis; PsO, psoriasis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

infections. Specifically, there were multiple spontaneous reports of tuberculosis among patients treated with infliximab [2]. Given the rarity of tuberculosis in the US, this relatively large case series (70 events) was unusual, and, extrapolating from estimates of the number of patients who had been treated with the drug up that point in time, the incidence of tuberculosis appeared higher than would have been expected. This prompted a black box warning from the US Food and Drug Administration (FDA) and a change in clinician behavior, with routine testing for prior tuberculosis exposure before initiation of treatment.

Subsequently, investigators have examined the association of biologics used in the treatment of immune-mediated diseases with the risk of other opportunistic infections and more common infections. For example, use of anti-TNF drugs has been associated with an increased risk of pneumonia among patients with inflammatory bowel disease [3]. Similar to tuberculosis, which can result from either *de novo* infection or reactivation of latent infection, there have been many cohort studies describing reactivation of hepatitis B among patients treated with anti-TNF therapy [4]. Numerous systematic reviews and meta-analyses have addressed this question. A 2016 systematic review of clinical trials of anti-TNF therapies for RA, ankylosing spondylitis, or psoriatic arthritis reported that anti-TNF therapy was associated with a 20% increase in any infection and a 40% increase in serious infections. This risk translates into an absolute rate difference of approximately 1–2 per 100 patient-years. There was no reported increase in opportunistic infections other than tuberculosis, but the authors noted that the rate of reporting of opportunistic infections in the clinical trials was very low [5].

Emphasizing the importance of the specificity of the targeting of different aspects of the immune system, two similar biologic drugs that target alpha-4 integrins have been marketed for the treatment of Crohn's disease. Natalizumab, which is also approved for the treatment of multiple sclerosis, targets α4β1 and α4β7; vedolizumab, which is also approved for the treatment of ulcerative colitis, targets only α4β7. Because α4β7 only directs lymphocytes to the gut, whereas α4β1 directs homing of lymphocytes to the gut and the brain, the risk of opportunistic infections between these two medications has proven to be quite different. Natalizumab has been associated with development of a typically severe, often fatal brain infection, progressive multifocal leukoencephalopathy (PML) [6]. However, this potentially fatal reactivation of JC virus infection has not been observed in patients treated with vedolizumab [7]. In the absence of immunosuppressant medication exposure, PML is typically only observed in the setting of other diseases associated with immunosuppression, such as HIV infection or malignancy. However, like most adverse drug reactions, PML is not uniquely associated with natalizumab and has been observed with other biologic therapies, most notably rituximab in RA [8], although cases with anti-TNF therapy have also been reported [9].

Neoplasia

Perhaps due to the dreaded nature of the disease, the fear of cancer has proven to be one of the major factors that limit the use of anti-TNF medications for the treatment of immune-mediated diseases. Early after the marketing of anti-TNF drugs, concern was raised about whether these medications increased the risk of cancer, particularly lymphoma. In *post hoc* analyses from premarketing clinical trials, there was evidence of an increased incidence of lymphoma among patients with RA who were treated with anti-TNF medications compared to rates expected in the general population. This led to black box warnings for anti-TNF medications for the risk of lymphoma [10–12]. However, these estimates were not based on

comparisons to patients with comparably severe RA who likely also had an increased risk of lymphoma [13,14]. Indeed, patients with more active RA may be at appreciably higher risk for lymphoma [15], consistent with the mechanistic explanation related to a greater burden of systemic inflammation and chronic immune activation. This hypothesis remains under study [16], and its relevance to pharmacoepidemiology lies in the potential for confounding by disease activity and associated systemic inflammation. In addition, many patients who are treated with anti-TNF drugs may have been previously or simultaneously treated with other medications that increase the risk of lymphoma, such as thiopurines [17]. This is particularly true in patients with IBD. Ultimately, many further studies have been published on this topic, often coming to differing conclusions in part due to low statistical power [18–23]. A recent large study from Sweden suggested that anti-TNF therapy was associated with a small increased risk of lymphoma, and that concomitant therapy with thiopurines increased that risk further [24].

The risk of other cancers with biologic therapies for immune-mediated diseases seems to be small, if it is increased at all. There are some reports of a potential increased risk of melanoma with anti-TNF drugs, but this has not been universally reproduced [25,26]. Because the incidence of melanoma in this population is so low (less than 1 per 1000 patient-years), recent efforts have pooled data from a variety of data sources (e.g., multiple European registries) and have not confirmed this association [27]. As with lymphoma, all assessments of cancer risk need also to take into consideration the potential impact of other prior and concurrent immunosuppressant medications [28]. In patients with a history of prior cancer, these medications are generally avoided until there is evidence that the cancer has been cured in that patient. However, with the possible exception of nonmelanoma skin cancer [29], the available data suggest that use of anti-TNF medications is relatively safe once prior cancers are thought to be cured [30,31].

Studies of the association of newer biologic drugs, in classes other than anti-TNF, focusing on cancer have been relatively small, limiting the ability to draw strong conclusions [32,33].

Paradoxical Reactions

There are a few scenarios where treatment with biologic drugs results in a paradoxical reaction, meaning that the drug causes a syndrome similar to one that it is used to treat when being used to treat a different disease. The prototypical example of this is the development of a psoriaform rash in patients treated with anti-TNF drugs [34]. The rash often responds to similar medications as are used to treat psoriasis, and frequently recurs with switching between anti-TNF drugs. The exact mechanism that drives these paradoxical reactions is unknown.

Other Adverse Events of Interest with Biologic Drugs

There are a number of other adverse events that have drawn attention with the biologic drugs used to treat immune-mediated diseases. Tocilizumab, an interleukin-6 receptor antagonist used to treat RA, has been associated with bowel perforation, typically of the lower gastrointestinal (GI) tract [35,36]. This pattern is distinct from the more typical sites of GI perforation in RA patients that are associated with nonsteroidal anti-inflammatory drugs, which are in the upper GI tract. Numerically adverse changes in low-density lipoprotein (LDL) have also been observed with some biologic therapies, including anti-TNF agents [37], and particularly the anti-IL6R class of medications (e.g., tocilizumab) [38]. While the ratio of LDL to high-density lipoprotein is unchanged, the potentially adverse changes in lipid profiles initially raised safety concerns for an increased risk

of cardiovascular disease (CVD) related events. Reassuringly, CVD event rates from the tocilizumab clinical trial program [39] and recent results from a large safety trial of more than 3000 patients with a CVD endpoint [40] showed no difference between anti-TNF therapy and tocilizumab. These results were confirmed by several observational analyses [41, 42, 43] that also suggested tocilizumab does not increase CVD risk.

The anti-TNF drugs frequently result in a positive antinuclear antibody and occasionally cause drug-induced lupus [44]. Notably, reports of severe liver injury and pancreatitis with the biologic drugs for immune-mediated disease have been relatively infrequent [45,46]. There have also been cases of new-onset IBD or exacerbations of preexisting IBD associated with anti-IL17 therapies, used for psoriasis and psoriatic arthritis, although a more systematic evaluation has not confirmed this association [47].

In contrast to the biologic drugs that are designed to suppress the immune system, the immune checkpoint inhibitors, such as ipilimumab and nivolumab, which are used as treatment for cancer, are designed to increase activity of the immune system. The prototypical adverse reaction to these agents is the onset of colitis that mimics IBD [48].

Finally, as might be expected with a complex molecule, hypersensitivity reactions to administration of biologics are a potential safety concern. These range from infusion reactions of mild severity that do not preclude continued treatment (generally with concomitant diphenhydramine and intravenous glucocorticoids), to anaphylactic reactions that may be life threatening and result in death. Although a few medications have associated biologic markers that suggest that the risk of hypersensitivity reactions may be increased (e.g., increasing serum uric acid for patients treated with pegloticase) or assays that can detect antidrug antibodies (ADAs; discussed shortly), some hypersensitivity reactions are idiosyncratic and unpredictable. This type of rare but often serious safety event may be particularly suitable to study using pharmacoepidemiologic methods (e.g., self-controlled study designs, see Chapter 43) [49], given the rather predictable temporal association between exposure and outcome. For example, anaphylactic-type reactions would be expected within 24 hours of intravenous administration.

Comparative Effectiveness of Biologic Drugs

Because many biologic drugs are foreign proteins, many have been routinely administered in combination with other immunosuppressant medications, with the hope of reducing the incidence of the development of ADAs and improving both efficacy and durability of response [50,51]. For most disease indications, biologic drugs have been shown to have enhanced effectiveness if given as combination therapy with a background conventional therapy (e.g., TNFi plus either methotrexate or thiopurines), as demonstrated in high-quality randomized controlled trials [52,53]. In contrast, a few trials have attempted to combine multiple biologics together. These have generally shown negligible incremental clinical benefit and a higher rate of adverse events, especially infections [54,55]. For the time being, it is therefore reasonable to assume for the purpose of characterizing exposure in a pharmacoepidemiologic study that patients are not simultaneously receiving two biologic medications for the same disease indication.

While being on two immunosuppressive or immunomodulatory treatments may have synergistic effects with respect to clinical effectiveness, use of the combination may affect safety. However, designing pharmacoepidemiology studies to compare combination therapy to monotherapy presents a unique challenge. Some patients will "step up" from monotherapy with one of the two medications to combination therapy,

while others will switch from one therapy to the other, and yet a third option is to start both medications around the same time, although generally not on the same day. Determining which treatment pattern applies to which patient can be difficult when using administrative claims data or even electronic health records. One often needs to observe several months of treatment to determine whether the first medication has been continued, in order to know whether the patient added the second medication and is now on two treatments, or switched to the second medication as monotherapy. To avoid immortal time bias, one approach is to start follow-up for all groups several months after the second drug was added to allow for clarity over to which treatment group each patient should be assigned [56]. Unfortunately, this approach limits the ability to study outcomes that occur shortly after the initiation of combination therapy.

The reasons that patients may not continue the background treatment (e.g., methotrexate) may also be important when studying outcomes. Some patients may discontinue the background therapy due to fear of future side effects in the setting of lack of perceived benefit when taken as monotherapy. Some patients may not continue treatment due to intolerance or other "nuisance" side effects (e.g., alopecia, oral ulcers) that have minimal impact on major health outcomes. Other patients may not continue background conventional treatment because of co-morbidities (e.g., chronic liver disease) or nonadherence to the required laboratory monitoring (e.g., liver enzyme testing). Thus, the underlying reasons why patients are not receiving combination therapy may be important from a pharmacoepidemiologic perspective, and may affect the risk for other types of adverse events. Measuring and controlling for these factors underlying use of monotherapy (versus combination therapy) are likely to be important to control for in any pharmacoepidemiologic study if these reasons (or proxies for them) are available in the analytic data source.

For most biologic indications, the large, Phase III studies conducted for regulatory purposes have compared biologics against placebo, and randomized comparative effectiveness trials are infrequent. In RA and IBD, for example, only a small handful of direct comparisons between biologics are available that are powered for superiority [57]; others have been designed as noninferiority studies against anti-TNF therapy [58]. Anti-TNF therapy is commonly used as the referent comparator agent, given that it was the first biologic class of drugs approved for most autoimmune and inflammatory conditions and has the greatest uptake worldwide. Indirect comparisons and meta-analyses have suggested relatively comparable efficacy between therapeutic agents, although changing from first anti-TNF therapy to a biologic with a different MOA may be preferable for patients with inadequate clinical efficacy [59]. Unfortunately, "inadequate clinical efficacy" and "loss of response" are imprecisely defined clinically and vary greatly by disease. More recently, newer biologics with different MOAs have been compared against anti-TNF therapy. Results from randomized controlled trials in psoriasis suggest that biologics targeting the IL-12/23 [60] and IL-17 pathways likely have superior efficacy in psoriasis compared to TNFi therapy [61]. IL-17 therapy appears to have superior efficacy even to ustekinumab [62]. However, this effect does not appear to extrapolate to other conditions, even those that are closely related (e.g., psoriatic arthritis, PsA).

While clinical effectiveness of improvement in inflammatory arthritis (RA, PsA), in gut inflammation (IBD), or in the skin (psoriasis) are endpoints that are difficult to assess in claims or electronic health record (EHR) data sources used for pharmacologic research, a number of attempts have been made to create and validate algorithms to serve as proxies for clinical endpoints that can be applied to administrative data. In RA, for example, an algorithm that requires patients to remain on the biologic

with good adherence, not add a new conventional therapy, undergo intra-articular injection or increase systemic glucocorticoids, has been shown to mirror the clinical effectiveness of RA treatments with good sensitivity, specificity, and positive predictive value [63]. For other biologic indications, the relevant outcome measures may be directly available in claims or EHR data sources. For example, effectiveness outcomes such as avoidance of bowel surgery (IBD) [56] or fracture (osteoporosis) [64,65] can be studied directly with good validity in these data sources.

Safety of Switching and Restarting

Unlike with traditional small-molecule drugs, a unique challenge of using biologic drugs is that each drug is seen by the patient's immune system as a foreign protein. Therefore, the patient's immune system is programmed to make ADAs directed against the biologic drug. These neutralizing antibodies lead to loss of response and increase the likelihood of having an allergic reaction to the medication.

Much has been learned about the risk of developing antibodies to biologics since the introduction of infliximab. When infliximab was initially approved for RA and Crohn's disease, it was used episodically, being administered when the patient became symptomatic again. This strategy allowed most patients to have sufficiently long gaps in therapy that the drug was completely cleared from the circulation. However, this would prime the immune system to recognize the drug at the next infusion and manifest with high levels of ADAs. The result was high rates of infusion reactions and loss of response. A landmark trial in Crohn's disease clearly demonstrated that induction therapy with three doses of medication followed by maintenance dosing was superior to on-demand dosing for maintenance of remission and prevention of ADAs [66].

In clinical practice, even when induction and maintenance dosing are employed, there are times when patients may need to have gaps in therapy or will desire to restart a biologic drug after a period off medications. Such gaps put the patient at risk for allergic reactions driven by ADAs, although in the era of maintenance dosing this seems to be less common [67]. Of importance is that the risk for ADAs is somewhat disease specific, such that patients with RA and IBD are at higher risk than patients with PsA, psoriasis, and ankylosing spondylitis. Regardless of the risk, concomitant therapy with methotrexate or thiopurines appears to significantly attenuate the risk of ADA formation, although does not abrogate it completely [68–71].

Switching drugs within the same class of biologics is common when considering the anti-TNF drugs, where there are five approved for a variety of indications. The most common indication for switching within class is loss of response to a drug within the same class. At least in IBD, measuring drug levels and antidrug antibodies can help guide this strategy. When there are high levels of ADAs and low or absent drug levels, this suggests that the loss of response was mediated by the ADAs and that the patient would be more likely to respond to another drug within the same class [72]. In contrast, if there are high drug levels and low or absent antidrug antibodies, switching to a drug of a different class is more likely to be effective [72]. At this point, measuring drug levels and/or ADAs is unique to the management of IBD. However, the relevance of this strategy may increase given the expected future availability of multiple biosimilar anti-TNF medications that could increase use of this class of medications relative to other biologics with different MOAs that do not have a biosimilar equivalent available.

There are circumstances where patients may choose to switch drugs within the same class for convenience. For example, patients may switch from an intravenous drug to a self-injected drug to avoid the need to schedule infusion visits. There are relatively little data on the safety of

such switches. A small trial in patients with Crohn's disease suggested that patients who were in remission on infliximab were more likely to stay in remission and without adverse events if they remained on infliximab than if they switched to adalimumab [73]. However, even though biologics may have the same mechanism of action and therapeutic target (e.g., TNF), it should not be assumed that the clinical response to one agent in the class will necessarily mirror the clinical response to another.

An evolving area of research is the question of switching between the originator drug and the biosimilar of the same drug. There are multiple definitions of biosimilars, but all are generally consistent with the FDA's definition that the compound is highly similar to the reference product, and that there are no clinically meaningful differences from the reference product in terms of safety, purity, and potency [74]. Approval of biosimilars for marketing has generally relied on a single clinical study showing noninferiority or equivalence, in which one group receives the originator drug and the other group the biosimilar. The potential cost savings of biosimilars are appreciable, up to 40–70% in Europe [75], although the savings are likely to be appreciably less in the US (e.g., 20–30%). There is interest in knowing whether one can safely switch between originator and biosimilar drugs in patients who are currently receiving therapy with the originator drug. The strongest evidence of the safety of this strategy comes from the NOR-Switch trial, in which patients with RA, Crohn's disease, ulcerative colitis, psoriasis, PsA, and spondyloarthritis who were stable on treatment with infliximab for at least 6 months were randomly assigned to continue the originator infliximab or switch to CT-P13 (infliximab-dyyb, Inflectra®). The primary outcome was disease worsening. The overall adjusted rate of worsening slightly favored the originator drug, but was not significant, and it met the predefined 15% threshold for equivalence (−4.4%, −12.7% to 3.9%) [76]. Although reassuring, there remain additional questions, such as the safety of repeated switching and switching between multiple biosimilars with respect to immunogenicity and the formation of ADAs.

Currently Available Solutions: Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

Accurate Identification of Biologic Exposure in Pharmacoepidemiology Data Sources

By their nature, biologic agents must be administered parenterally, either subcutaneously or by intravenous administration. Some therapies (e.g., abatacept, tocilizumab, golimumab) can be given via either route of administration, and ustekinumab, when used for Crohn's disease, is given intravenously for the first dose and then subcutaneously thereafter. For some biologics, clinicians can vary either the dose and/or the frequency of administration. Dosing variation generally takes one of three forms, specific to each biologic and its indication(s): (i) fixed dose and frequency (e.g., Etanercept 50 mg once weekly for RA); (ii) fixed dose, with the possibility of increased frequency depending on clinical response (e.g., adalimumab 40 mg every 2 weeks, with possibility of increase to 40 mg once weekly); and (iii) weight-based dosing (e.g., infliximab 5 mg/kg every 8 weeks, with possibility to both increase the dose and increase the frequency depending on clinical response). Self-injected drugs are reimbursed under a pharmacy benefit (e.g., Medicare part D), and intravenous treatments are reimbursed under a medical benefit (e.g., Medicare part B). Infused biologics will be identifiable in most administrative data sources using Healthcare Common Procedure Code System (HCPCS) Level II codes as claims for medical procedures (e.g., infliximab, J1745). A few biologics can be reimbursed under either

the pharmacy benefit (e.g., certolizumab SQ, self-administered) or under the medical benefit (certolizumab SQ, administered once monthly by a healthcare provider).

Biologics identified in pharmacy data using National Drug Codes (NDC) are generally straightforward to classify in terms of the dose and quantity dispensed, given the specificity of NDC codes. Biologics billing as medical procedures will typically have dose represented based on the units appearing on the medical procedure claim. For example, the HCPCS code J1745 for infliximab represents 10 mg, so 30 units would reflect a dose of 300 mg. Given that the starting dose of this medication is generally fixed by disease indication, one can infer the patient's weight (or weight category) based on units dispensed. For example, in IBD, the starting dose is typically 5 mg/kg, so an initial dose of 300 mg, 400 mg, or 500 mg could be reasonably used to assume that the patient's weight was approximately 60 kg, 80 kg, or 100 kg, respectively. While not perfectly precise, given that the drug is provided in 100 mg vials and healthcare providers typically round up to the nearest whole vial, this approach is nevertheless potentially helpful to infer the patient's weight category in data sources where it is not directly available (e.g., health plan claims).

One challenge in identifying biologics given as medical procedures is that after a biologic is approved, it is assigned a "nonspecific" HCPCS code by the Center for Medicare and Medicaid Services. This is typically replaced during the next year (or two) by a permanent HCPCS code that uniquely identifies the drug. Until this occurs, however, the drug is reimbursed under J3490 (Unclassified drug) or J3590 (Unclassified biologic). A combination of the associated disease indication (using International Classification of Diseases, ICD) codes, units dispensed, submitted or allowed amount, and other features associated with the procedure claim can be used to specifically identify the drug [77]. The relevance for pharmacoepidemiology studies is that if this approach is not followed, it will fail to identify the earliest users of the medication, and it will misclassify the date of first use. The implications are that any new user study design [78] of biologics in their first 1–2 years after licensure may be meaningfully compromised if this issue is not appropriately addressed methodologically.

An additional consideration is to understand the duration of exposure in relation to each biologic administration. As with other therapies, the effect of biologic drugs reflects not only the half-life of the agents (i.e., pharmacokinetics), but also the duration of their effect on the body (i.e., pharmacodynamics). Subject matter expertise is therefore needed to understand whether an "extension" to current exposure might be warranted. This decision depends on the drug, its pharmacokinetics and pharmacodynamics, and the outcome being studied. For outcomes with a short expected interval between exposure and outcome, a relatively short (e.g., 30–90-day extension) is likely warranted [79]. As with all pharmacoepidemiology studies, researchers also must consider whether the relative hazard of the outcome is likely to be constant or time varying. For example, the association between biologic exposure and serious infections has been shown to peak early and then flatten [80]. A variety of factors may account for this, including a true biologic effect reflecting a reduction in systemic inflammation (which has been associated with infection risk), subsequent reduction in glucocorticoid use, or depletion of susceptible patients [81]. In contrast, for an outcome like malignancy, where the shape of the hazard curve between exposure and outcome is not clear, "ever exposed" may be preferable, and risk with increasing cumulative exposure in discrete time intervals should be estimated over the follow-up period. Because switching of therapies is common, particularly when the first medication fails to achieve complete symptom control, all of these studies must also consider the biologic plausibility of very early outcomes and those that occur long after a medication is discontinued, and the potential impact of prior, concomitant, and subsequent treatments on the outcome of interest.

Biosimilar Naming Conventions and Representation in Administrative Data Sources

With generic drugs, the generic product is referred to solely by the name of the generic compound, since the medication is an exact replicate of the active ingredient. Biosimilars are not exact replicates of the referent biologic. Therefore, it is necessary for each biosimilar to have a unique name, but still have a name that reflects the nature of the product. Based on this, the FDA has proposed that biologics all have a four-letter suffix added to the generic name. By requirement, this suffix can have no clinical meaning. For example, the original infliximab (Remicade®) has the suffix -hjmt (although current guidance on nomenclature suggests to omit the suffix for the reference biologic product), while the biosimilar Inflectra® is infliximab-dyyb. These unique suffixes will allow for accurate identification of the biosimilars within claims data and electronic health records [82]. Biosimilars administered by healthcare facilities will continue to be identified by HCPCS codes in the US. Prior to January 1, 2018, the HCPCS codes for biosimilars to the same originator product were the same, but with a unique modifier that would be included in the billing data. For example, Inflectra® and Renflexis® share the same HCPCS code (Q5102) but have unique modifiers (ZB and ZC, respectively) [83]. However, when biosimilars are first released, before an HCPCS code is assigned, they may need to be identified by a generic J code, similar to the challenges of identification of any new biologic administered by a healthcare facility. After January 1, 2018, these rules have changed and guidance is now available from the Centers for Medicare and Medicaid Services [83]. Drug-specific codes for individual biosimilars are now assigned. For example, new medical procedure codes were created to uniquely identify the infliximab biosimilars, Q5103 (Inflectra®) and Q5104 (Renflexis®). A relatively similar system has been proposed for the European Union, but complete harmonization has not yet occurred and seems highly desirable [82].

Confounding by Indication

As with all pharmacoepidemiology and comparative effectiveness studies, confounding by indication is an important consideration. This may be an even greater concern with studies of biologic medications, due to the high cost of these medications leading to restrictions on access by some insurance plans until patients have failed to adequately respond to other less expensive medications or through other means [84,85]. The methods to address confounding by indication (see Chapter 43) in studies of biologics do not differ from those of other medications, but should take into consideration this channeling process. Patients who are being treated with a second or third biologic are also likely to be more refractory to therapy than those receiving their first biologic drug. For example, in patients with IBD, those receiving vedolizumab were much more likely to respond to therapy if they had not been previously treated with an anti-TNF drug [86,87]. The reasons for switching from a first to a second (or subsequent) biologic are also likely important. For example, if patients with RA have had an inadequate clinical response to the first TNF drug, they are more likely to have an inadequate response to a subsequent TNF drug [88]. However, if they discontinued the first TNF drug for a safety or tolerability concern, then the clinical response to the subsequent TNF drug was unaffected, but it is more likely to fail for safety or tolerability reasons. In most administrative data sources, the reasons for discontinuing or switching therapy are not known, but some registries or EHR data sources may capture this information.

More refractory patients who have received multiple biologics in the past also will typically have longer disease duration, a higher comorbidity burden, and may be at higher risk for serious adverse events such as infections [89].

Thus, having sufficient historical data to fully account for the number of prior medications (both biologic and conventional therapies) with which the patient has been treated, particularly the number of different biologic and conventional medications that the patient has received, is important to fully capture the disease severity. Ascertainment of prior treatment history is likely to be optimal if an extended baseline period of observation is available, and a variable "baseline" period (e.g., 12 months, plus all available prior data) may be useful for assessment of these specific covariates [90,91].

To overcome confounding by indication, a number of considerations relevant to biologic access may be useful. Because health plans often dictate the selection of which biologic must be used first (and sometimes second), in what are called "fail-first" policies, then selection is not predicated on individual patient characteristics, and patients with the same condition treated with different first-line biologic agents may be compared between health plans. Assuming the health plan does not also dictate which second-line biologic treatment the patient may receive, there are major evidence gaps with respect to which biologic should be optimally be used next, and physician judgment (in the absence of evidence) may prevail. Instrumental variable methods (Chapter 43) using provider preference (which may be affected by financial motivations to use infusion-based treatments [92]) or other logistics related to biologic use (e.g., driving distance to the infusion center for once-monthly infusion treatments) may be useful methodologic considerations to address confounding by indication and reduce bias.

Measuring Treatment Discontinuation

A related concept is measuring treatment discontinuation. Persistence on therapy is a commonly used endpoint in comparative effectiveness studies that is used as a surrogate for sufficient response to therapy to justify continued use of these expensive immunosuppressive therapies [56,93–95]. Similarly, assigning an exposure category to an event requires knowledge of whether the patient was exposed to the therapy when the event occurred. Because many biologic drugs are dosed intermittently due to long half-lives, defining discontinuation can take several months after the last dispensing to determine whether the patient discontinued the medication. Furthermore, because many biologic drugs suppress the immune system, clinicians may temporarily hold these medications around the time of elective surgeries or if there is evidence of an infection. Thus, it is important to account for potential gaps when considering persistence on therapy. Defining the optimal duration of "gap" time while considering a patient to have continuous therapy as opposed to reinitiation of therapy can sometimes be determined empirically from the available data by examining the pattern of drug dispensing, particularly focusing on dispensing that appears to be a reinitiation regimen based on the dose. Finally, as noted previously, because patients who have lost response to other biologics are more likely to lose response to the next biologic, accounting for prior biologic use is important in studies examining persistence on therapy.

The Future

New Therapies

Although biologics have proved very effective for the treatment of autoimmune and inflammatory diseases, new small-molecule targeted therapies have emerged that have similarly rapid onset (e.g., demonstrable benefit in the first 2 weeks of treatment and near-maximal effect by 12 weeks) and comparable effectiveness and safety to biologics. Unlike conventional and slower-acting immunosuppressive drugs for these same indications that have may have multiple biologic effects, new synthetic immunomodulatory drugs that target specific immune and inflammatory pathways (e.g., the JAK/STAT signaling pathway)

have emerged and proven very effective compared to biologics [96]. Some are already approved in the US (e.g., tofacitinib for RA, PsA, and UC; baricitinib for RA) and Europe (e.g., baricitinib for RA) and have particular safety concerns based on their mechanisms of action (e.g., increased rate of herpes zoster observed with janus kinase inhibitors, likely as a consequence of effects on interferon gamma, and possibly increased rates of venous thromboembolism when JAK/STAT drugs are used at higher doses) [97]. Several other compounds targeting the JAK/STAT pathway, as well as other immune-related pathways (e.g., spleen tyrosine kinase, Syk) are in development and will continue to spur interest in real-world evidence comparing these therapies to current and future biologics.

New Data Sources and Linked Data to Generate Real-World Evidence

To date, many studies of biologic safety have been conducted in large administrative healthcare databases, and disease-specific or drug-specific (biologic) registries [98–101]. Recognizing the limitations of each of these types of data, there has been growing interest in expanding both the kinds of data sources used for comparative evaluation of biologic safety and effectiveness, as well as linking between data sources. Large-scale EHR data such as that available in the Patient-Centered Clinical Research Network (PCORnet) [102–104] and from specialty-specific registries such as the American College of Rheumatology's (ACR) Rheumatology Informatics System for Effectiveness (RISE) [105] offer the potential for large-scale data analysis on patients treated with biologics. EHR data will generally provide more precise phenotypic information (including laboratory results) than that available in health plan claims data alone. Beyond the many traditional registries available for the study of biologic safety and effectiveness, a growing number of patient-focused research registries in IBD, RA, psoriasis and PsA, multiple sclerosis, and juvenile idiopathic arthritis [106,107] have also been established and may provide unique information from a patient's perspective. Many of these capture information on biologic use and outcomes using validated patient-reported outcome (PRO) instruments, such as those available from the National Institute of Health's Patient-Reported Outcome Management System (PROMIS) [108].

A variety of methods may be used to link administrative claims, EHR, and traditional registry data together, and combine them with other data (e.g., disease-specific biomarker data). Deterministic methods applied to unique identifiers (e.g., social security numbers or health plan identifiers) can be used for linkage, but doing so may invoke privacy-related concerns. Probabilistic linkage using multiple, non-unique identifiers (e.g., patient date of birth, visit dates, national provider identifiers, number of a patient's physician, dates of outpatient clinic visits, or admission/discharge dates of hospitalization (for hospitalized patients, or those undergoing in-hospital procedures) has been used successfully to link patients [109]. When no identifiers can be shared, a variety of privacy-preserving methods are also available (e.g., one-way hashing algorithms) to allow for sharing of minimally identifiable information, fully encrypted data, or statistical coefficients [110–112]. A number of use cases that have linked clinical information (e.g., disease activity) to registries or lab-based data sources with outcomes commonly available in health plan data (e.g., healthcare costs, all-cause hospitalization) have been published in biologic-exposed patients [113,114].

Value-Based Care

Given the high cost of biologics, often in the range of several thousands of dollars per patient per month or more, increasing attention is placed on demonstrating and maximizing the value of these therapies. While value-based reimbursement, risk-sharing contracts, and so on are by no means unique to the use of biologic agents, their high cost and the need to

characterize the safety profile of new molecular entities compared to existing treatments places great importance on the careful study of these medications. Pharmacoepidemiologic research is therefore likely to continue to play a prominent role in filling evidence gaps to demonstrate improved clinical and patient-focused outcomes, and to promote the rational use of biologic therapies in the form of evidence-based guidelines.

References

1 Rodgers KR, Chou RC. Therapeutic monoclonal antibodies and derivatives: historical perspectives and future directions. *Biotechnol Adv* 2016; **34**: 1149–58.

2 Keane J, Gershon S, Wise RP, *et al*. Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med* 2001; **345**: 1098–104.

3 Long MD, Martin C, Sandler RS, *et al*. Increased risk of pneumonia among patients with inflammatory bowel disease. *Am J Gastroenterol* 2013; **108**: 240–8.

4 Moghoofei M, Mostafaei S, Ashraf-Ganjouei A, *et al*. HBV reactivation in rheumatic diseases patients under therapy: a meta-analysis. *Microb Pathog* 2017; **114**: 436–43.

5 Minozzi S, Bonovas S, Lytras T, *et al*. Risk of infections using anti-TNF agents in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: a systematic review and meta-analysis. *Expert Opin Drug Saf* 2016; **15**: 11–34.

6 Rudick RA, O'Connor PW, Polman CH, *et al*. Assessment of JC virus DNA in blood and urine from natalizumab-treated patients. *Ann Neurol* 2010; **68**: 304–10.

7 Bye WA, Jairath V, Travis SPL. Systematic review: the safety of vedolizumab for the treatment of inflammatory bowel disease. *Aliment Pharmacol Ther* 2017; **46**: 3–15.

8 Bharat A, Xie F, Baddley JW, *et al*. Incidence and risk factors for progressive multifocal leukoencephalopathy among patients with selected rheumatic diseases. *Arthritis Care Res (Hoboken)* 2012; **64**: 612–15.

9 Ray M, Curtis JR, Baddley JW. A case report of progressive multifocal leucoencephalopathy (PML) associated with adalimumab. *Ann Rheum Dis* 2014; **73**: 1429–30.

10 Remicade Prescribing Information, 2013. www.remicade.com.

11 Humira Prescribing Information, 2017. www.humira.com.

12 Etanercept Prescribing Information, 2017. www.enbrel.com.

13 Baecklund E, Ekbom A, Sparen P, *et al*. Disease activity and risk of lymphoma in patients with rheumatoid arthritis: nested case-control study. *BMJ* 1998; **317**: 180–1.

14 Baecklund E, Iliadou A, Askling J, *et al*. Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**: 692–701 [see comment].

15 Baecklund E, Iliadou A, Askling J, *et al*. Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**: 692–701.

16 Hellgren K, Baecklund E, Backlin C, *et al*. Rheumatoid arthritis and risk of malignant lymphoma: is the risk still increased? *Arthritis Rheumatol* 2017; **69**: 700–8.

17 Kotlyar DS, Lewis JD, Beaugerie L, *et al*. Risk of lymphoma in patients with inflammatory bowel disease treated with azathioprine and 6-mercaptopurine: a meta-analysis. *Clin Gastroenterol Hepatol* 2015; **13**: 847–58 e4; quiz e48–50.

18 Askling J, Fored CM, Baecklund E, *et al*. Haematopoietic malignancies in rheumatoid arthritis: lymphoma risk and characteristics after exposure to tumour necrosis factor

antagonists. *Ann Rheum Dis* 2005; **64**: 1414–20.

19 Brown SL, Greene MH, Gershon SK, *et al.* Tumor necrosis factor antagonist therapy and lymphoma development: twenty-six cases reported to the Food and Drug Administration. *Arthritis Rheum* 2002; **46**: 3151–8.

20 Herrinton LJ, Liu L, Weng X, *et al.* Role of thiopurine and anti-TNF therapy in lymphoma in inflammatory bowel disease. *Am J Gastroenterol* 2011; **106**: 2146–53.

21 Siegel C, Marden S, Persing S, *et al.* Risk of lymphoma associated with anti-TNF agents for the treatment of Crohn's disease: a meta-analysis. *Gastroenterol* 2009; **7**(8): 874–81.

22 Wolfe F, Michaud K. The effect of methotrexate and anti-tumor necrosis factor therapy on the risk of lymphoma in rheumatoid arthritis in 19,562 patients during 89,710 person-years of observation. *Arthritis Rheum* 2007; **56**: 1433–9.

23 Siegel CA, Marden SM, Persing SM, *et al.* Risk of lymphoma associated with combination anti-tumor necrosis factor and immunomodulator therapy for the treatment of Crohn's disease: a meta-analysis. *Clin Gastroenterol Hepatol* 2009; **7**: 874–81.

24 Lemaitre M, Kirchgesner J, Rudnichi A, *et al.* Association between use of thiopurines or tumor necrosis factor antagonists alone or in combination and risk of lymphoma in patients with inflammatory bowel disease. *JAMA* 2017; **318**: 1679–86.

25 Long MD, Martin CF, Pipkin CA, *et al.* Risk of melanoma and nonmelanoma skin cancer among patients with inflammatory bowel disease. *Gastroenterology* 2012; **143**: 390–9 e1.

26 Wolfe F, Michaud K. Biologic treatment of rheumatoid arthritis and the risk of malignancy: analyses from a large US observational study. *Arthritis Rheum* 2007; **56**: 2886–95.

27 Mercer LK, Askling J, Raaschou P, *et al.* Risk of invasive melanoma in patients with rheumatoid arthritis treated with biologics: results from a collaborative project of 11 European biologic registers. *Ann Rheum Dis* 2017; **76**: 386–91.

28 Osterman MT, Sandborn WJ, Colombel JF, *et al.* Increased risk of malignancy with adalimumab combination therapy, compared with monotherapy, for Crohn's disease. *Gastroenterology* 2014; **146**: 941–9.

29 Scott FI, Mamtani R, Brensinger CM, *et al.* Risk of nonmelanoma skin cancer associated with the use of immunosuppressant and biologic agents in patients with a history of autoimmune disease and nonmelanoma skin cancer. *JAMA Dermatol* 2016; **152**: 164–72.

30 Mamtani R, Clark AS, Scott FI, *et al.* Association between breast cancer recurrence with immunosuppression in rheumatoid arthritis and inflammatory bowel disease: a cohort study. *Arthritis Rheumatol* 2016; **68**(10): 2403–11.

31 Shelton E, Laharie D, Scott FI, *et al.* Cancer recurrence following immune-suppressive therapies in patients with immune-mediated diseases: a systematic review and meta-analysis. *Gastroenterology* 2016; **151**: 97–109 e4.

32 Simon TA, Smitten AL, Franklin J, *et al.* Malignancies in the rheumatoid arthritis abatacept clinical development programme: an epidemiological assessment. *Ann Rheum Dis* 2009; **68**: 1819–26.

33 Solomon DH, Kremer JM, Fisher M, *et al.* Comparative cancer risk associated with methotrexate, other non-biologic and biologic disease-modifying anti-rheumatic drugs. *Sem Arthritis Rheum* 2014; **43**: 489–97.

34 Cleynen I, Vermeire S. Paradoxical inflammation induced by anti-TNF agents in patients with IBD. *Nat Rev Gastroenterol Hepatol* 2012; **9**: 496–503.

35 Schiff MH, Kremer JM, Jahreis A, *et al.* Integrated safety in tocilizumab clinical trials. *Arthritis Res Ther* 2011; **13**: R141.

36 Xie F, Yun H, Bernatsky S, *et al*. Brief report: risk of gastrointestinal perforation among rheumatoid arthritis patients receiving tofacitinib, tocilizumab, or other biologic treatments. *Arthritis Rheumatol* 2016; **68**: 2612–17.

37 Daien CI, Duny Y, Barnetche T, *et al*. Effect of TNF inhibitors on lipid profile in rheumatoid arthritis: a systematic review with meta-analysis. *Ann Rheum Dis* 2012; **71**: 862–8.

38 Curtis JR, Xie F, Chen L, *et al*. The incidence of gastrointestinal perforations among rheumatoid arthritis patients. *Arthritis Rheum* 2011; **63**: 346–51.

39 Rao VU, Pavlov A, Klearman M, *et al*. An evaluation of risk factors for major adverse cardiovascular events during tocilizumab therapy. *Arthritis Rheumatol* 2015; **67**: 372–80.

40 Giles JT, Sattar N, Gabriel SE, *et al*. Comparative cardiovascular safety of tocilizumab vs etanercept in rheumatoid arthritis: results of a randomized, parallel-group, multicenter, noninferiority, Phase 4 clinical trial [abstract]. *Arthritis Rheumatol* 2016; **68**(suppl 10). https://acrabstracts.org/abstract/comparative-cardiovascular-safety-of-tocilizumab-vs-etanercept-in-rheumatoid-arthritis-results-of-a-randomized-parallel-group-multicenter-noninferiority-phase-4-clinical-trial (accessed May 2019).

41 Zhang J, Xie F, Yun H, *et al*. Comparative effects of biologics on cardiovascular risk among older patients with rheumatoid arthritis. *Ann Rheum Dis* 2016; **75**: 1813–18.

42 Xie F, Yun H, Levitan EB, Muntner P, Curtis JR. Tocilizumab and the risk of cardiovascular disease: direct comparison among biologic disease-modifying antirheumatic drugs for rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)* 2019 (in press) doi: 10.1002/acr.23737

43 Kim SC, Solomon DH, Rogers JR, *et al*. Cardiovascular safety of tocilizumab versus tumor necrosis factor inhibitors in patients with rheumatoid arthritis: a multi-database cohort study. *Arthritis Rheumatol* 2017; **69**: 1154–64.

44 Perez-De-Lis M, Retamozo S, Flores-Chavez A, *et al*. Autoimmune diseases induced by biological agents: a review of 12,731 cases (BIOGEAS Registry). *Expert Opin Drug Saf* 2017; **16**: 1255–71.

45 Colombel JF, Sands BE, Rutgeerts P, *et al*. The safety of vedolizumab for ulcerative colitis and Crohn's disease. *Gut* 2017; **66**: 839–51.

46 Shelton E, Chaudrey K, Sauk J, *et al*. New onset idiosyncratic liver enzyme elevations with biological therapy in inflammatory bowel disease. *Aliment Pharmacol Ther* 2015; **41**: 972–9.

47 Deodhar AA, Schreiber S, Gandh iK, *et al*. No increased risk of inflammatory bowel disease among secukinumab-treated patients with moderate to severe psoriasis, psoriatic arthritis, or ankylosing spondylitis: data from 14 phase 2 and phase 3 clinical studies [abstract]. *Arthritis Rheumatol* 2016; **68**(suppl 10). https://acrabstracts.org/abstract/no-increased-risk-of-inflammatory-bowel-disease-among-secukinumab-treated-patients-with-moderate-to-severe-psoriasis-psoriatic-arthritis-or-ankylosing-spondylitis-data-from-14-phase-2-and-phase-3-c (accessed May 2019).

48 De Velasco G, Je Y, Bosse D, *et al*. Comprehensive meta-analysis of key immune-related adverse events from CTLA-4 and PD-1/PD-L1 inhibitors in cancer patients. *Cancer Immunol Res* 2017; **5**: 312–18.

49 Yun H, Xie F, Beyl RN, *et al.* Risk of hypersensitivity to biologic agents among medicare patients with rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2017; **69**: 1526–34.

50 Singh JA, Saag KG, Bridges SL, Jr., *et al*. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol* 2016; **68**: 1–26.

51 Terdiman JP, Gruss CB, Heidelbaugh JJ, *et al*. American Gastroenterological Association Institute guideline on the use of thiopurines, methotrexate, and anti-TNF-alpha biologic

drugs for the induction and maintenance of remission in inflammatory Crohn's disease. *Gastroenterology* 2013; **145**: 1459–63.

52 Singh JA, Hossain A, Tanjong Ghogomu E, *et al*. Biologics or tofacitinib for rheumatoid arthritis in incomplete responders to methotrexate or other traditional disease-modifying anti-rheumatic drugs: a systematic review and network meta-analysis. *Cochrane Database Syst Rev* 2016; **5**: CD012183. doi:10.1002/14651858.CD012183

53 Colombel JF, Sandborn WJ, Reinisch W, *et al*. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med* 2010; **362**: 1383–95.

54 Genovese MC, Cohen S, Moreland L, *et al*. Combination therapy with etanercept and anakinra in the treatment of patients with rheumatoid arthritis who have been treated unsuccessfully with methotrexate. *Arthritis Rheum* 2004; **50**: 1412–19.

55 Weinblatt M, Schiff M, Goldman A, *et al*. Selective costimulation modulation using abatacept in patients with active rheumatoid arthritis while receiving etanercept: a randomised clinical trial. *Ann Rheum Dis* 2007; **66**: 228–34.

56 Osterman MT, Haynes K, Delzell E, *et al*. Effectiveness and safety of immunomodulators with anti-tumor necrosis factor therapy in Crohn's disease. *Clin Gastroenterol Hepatol* 2015; **13**: 1293–301 e5; quiz e70, e72.

57 Gabay C, Emery P, van Vollenhoven R, *et al.* Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): a randomised, double-blind, controlled phase 4 trial. *Lancet* 2013; **381**: 1541–50.

58 Weinblatt ME, Schiff M, Valente R, *et al*. Head-to-head comparison of subcutaneous abatacept versus adalimumab for rheumatoid arthritis. *Arthritis Rheum* 2013; **65**(1): 28–38.

59 Schoels M, Aletaha D, Smolen JS, *et al*. Comparative effectiveness and safety of biological treatment options after tumour necrosis factor alpha inhibitor failure in rheumatoid arthritis: systematic review and indirect pairwise meta-analysis. *Ann Rheum Dis* 2012; **71**: 1303–8.

60 Griffiths CE, Strober BE, van de Kerkhof P, *et al*. Comparison of ustekinumab and etanercept for moderate-to-severe psoriasis. *N Engl J Med* 2010; **362**: 118–28.

61 Langley RG, Elewski BE, Lebwohl M, *et al*. Secukinumab in plaque psoriasis: results of two phase 3 trials. *N Engl J Med* 2014; **371**: 326–38.

62 Blauvelt A, Reich K, Tsai TF, *et al*. Secukinumab is superior to ustekinumab in clearing skin of subjects with moderate-to-severe plaque psoriasis up to 1 year: results from the CLEAR study. *J Am Acad Dermatol* 2017; **76**: 60–69 e9.

63 Curtis JR, Baddley JW, Yang S, *et al*. Derivation and preliminary validation of an administrative claims-based algorithm for the effectiveness of medications for rheumatoid arthritis. *Arthritis Res Ther* 2011; **13**: R155.

64 Bernatsky S, Lix L, O'Donnell S, *et al*. Consensus statements for the use of administrative health data in rheumatic disease research and surveillance. *J Rheumatol* 2013; **40**: 66–73.

65 Wright NC, Daigle SG, Melton ME, *et al*. The design and validation of a new algorithm to identify incident fractures in administrative claims data. *J Bone Miner Res* 2019 (in press). doi: 10.1002/jbmr.3807

66 Rutgeerts P, Feagan BG, Lichtenstein GR, *et al*. Comparison of scheduled and episodic treatment strategies of infliximab in Crohn's disease. *Gastroenterology* 2004; **126**: 402–13 [see comment].

67 Louis E, Mary JY, Vernier-Massouille G, *et al*. Maintenance of remission among patients with Crohn's disease on antimetabolite therapy after infliximab therapy is stopped. *Gastroenterology* 2012; **142**: 63–70 e5; quiz e31.

68 Krieckaert CL, Nurmohamed MT, Wolbink GJ. Methotrexate reduces immunogenicity in

adalimumab treated rheumatoid arthritis patients in a dose dependent manner. *Ann Rheum Dis* 2012; **71**: 1914–15.

69 Jani M, Barton A, Warren RB, *et al*. The role of DMARDs in reducing the immunogenicity of TNF inhibitors in chronic inflammatory diseases. *Rheumatology (Oxford)* 2014; **53**: 213–22.

70 Bartelds GM, Krieckaert CL, Nurmohamed MT, *et al*. Development of antidrug antibodies against adalimumab and association with disease activity and treatment failure during long-term follow-up. *JAMA* 2011; **305**: 1460–8.

71 Bartelds GM, Wijbrandts CA, Nurmohamed MT, *et al*. Anti-infliximab and anti-adalimumab antibodies in relation to response to adalimumab in infliximab switchers and anti-TNF naive patients: a cohort study. *Ann Rheum Dis* 2010; **69**(5): 817–21.

72 Afif W, Loftus EV, Jr., Faubion WA, *et al*. Clinical utility of measuring infliximab and human anti-chimeric antibody concentrations in patients with inflammatory bowel disease. *Am J Gastroenterol* 2010; **105**: 1133–9.

73 Van Assche G, Vermeire S, Ballet V, *et al*. Switch to adalimumab in patients with Crohn's disease controlled by maintenance infliximab: prospective randomised SWITCH trial. *Gut* 2012; **61**: 229–34.

74 Food and Drug Administration. Biosimilar and interchangeable products. 2017. https://www.fda.gov/drugs/biosimilars/biosimilar-and-interchangeable-products (accessed May 2019).

75 Yu B. Greater potential cost savings with biosimilar use. *Am J Manag Care* 2016; **22**: 378.

76 Jorgensen KK, Olsen IC, Goll GL, *et al*. Switching from originator infliximab to biosimilar CT-P13 compared with maintained treatment with originator infliximab (NOR-SWITCH): a 52-week, randomised, double-blind, non-inferiority trial. *Lancet* 2017; **389**: 2304–16.

77 Curtis JR, Xie F, Chen R, *et al*. Identifying newly approved medications in Medicare claims data: a case study using tocilizumab. *Pharmacoepidemiol Drug Saf* 2013; **22**: 1214–21.

78 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**: 915–20.

79 Dixon WG, Symmons DP, Lunt M, *et al*. Serious infection following anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis: lessons from interpreting data from observational studies. *Arthritis Rheum* 2007; **56**: 2896–904.

80 Askling J, Fored CM, Brandt L, *et al*. Time-dependent increase in risk of hospitalisation with infection among Swedish RA patients treated with TNF antagonists. *Ann Rheum Dis* 2007; **66**: 1339–44.

81 Strangfeld A, Eveslage M, Schneider M, *et al*. Treatment benefit or survival of the fittest: what drives the time-dependent decrease in serious infection rates under TNF inhibition and what does this imply for the individual patient? *Ann Rheum Dis* 2011; **70**: 1914–20.

82 Danese S, Bonovas S, Peyrin-Biroulet L. Biosimilars in IBD: from theory to practice. *Nat Rev Gastroenterol Hepatol* 2017; **14**: 22–31.

83 Centers for Medicare and Medicaid Services. Part B Biosimilar Biological Product Payment and Required Modifiers. 2017. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Part-B-Drugs/McrPartBDrugAvgSalesPrice/Part-B-Biosimilar-Biological-Product-Payment.html (accessed May 2019).

84 Yazdany J, Dudley RA, Chen R, *et al*. Coverage for high-cost specialty drugs for rheumatoid arthritis in Medicare Part D. *Arthritis Rheumatol* 2015; **67**: 1474–80.

85 Yadav A, Foromera J, Feuerstein I, *et al*. Variations in health insurance policies regarding biologic therapy use in inflammatory bowel disease. *Inflamm Bowel Dis* 2017; **23**: 853–7.

86 Feagan BG, Rutgeerts P, Sands BE, *et al*. Vedolizumab as induction and maintenance

therapy for ulcerative colitis. *N Engl J Med* 2013; **369**: 699–710.

87 Sandborn WJ, Feagan BG, Rutgeerts P, *et al*. Vedolizumab as induction and maintenance therapy for Crohn's disease. *N Engl J Med* 2013; **369**: 711–21.

88 Hyrich KL, Lunt M, Watson KD, *et al*. Outcomes after switching from one anti-tumor necrosis factor alpha agent to a second anti-tumor necrosis factor alpha agent in patients with rheumatoid arthritis: results from a large UK national cohort study. *Arthritis Rheum* 2007; **56**: 13–20.

89 Curtis JR, Xie F, Chen L, *et al*. The comparative risk of serious infections among rheumatoid arthritis patients starting or switching biological agents. *Ann Rheum Dis* 2011; **70**: 1401–6.

90 Nakasian SS, Rassen JA, Franklin JM. Effects of expanding the look-back period to all available data in the assessment of covariates. *Pharmacoepidemiol Drug Saf* 2017; **26**: 890–9.

91 Colantonio LD, Kent ST, Huang L, *et al*. Algorithms to identify statin intolerance in Medicare administrative claim data. *Cardiovasc Drugs Ther* 2016; **30**: 525–33.

92 Zhang J, Xie F, Delzell E, *et al*. Trends in the use of biologic therapies among rheumatoid arthritis patients enrolled in the U.S. Medicare Program. *Arthritis Care Res (Hoboken)* 2013; **65**(11): 1743–51.

93 Osterman MT, Haynes K, Delzell E, *et al*. Comparative effectiveness of infliximab and adalimumab for Crohn's disease. *Clin Gastroenterol Hepatol* 2014; **12**: 811–17 e3.

94 Gendelman O, Weitzman D, Rosenberg V, *et al*. Characterization of adherence and persistence profile in a real-life population of patients treated with adalimumab. *Br J Clin Pharmacol* 2018; **84**(4): 786–95.

95 Sauer BC, Teng CC, Tang D, *et al*. Persistence with conventional triple therapy versus a tumor necrosis factor inhibitor and methotrexate in US veterans with rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2017; **69**: 313–22.

96 van Vollenhoven RF, Fleischmann R, Cohen S, *et al*. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. *N Engl J Med* 2012; **367**: 508–19.

97 Winthrop KL, Yamanaka H, Valdez H, *et al*. Herpes zoster and tofacitinib therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2014; **66**: 2675–84.

98 Kremer JM. The CORRONA database. *Autoimmun Rev* 2006; **5**: 46–54.

99 Zink A, Askling J, Dixon WG, *et al*. European biologics registers – methodology, selected results, and perspectives. *Ann Rheum Dis* 2009; **68**(8): 1240–6.

100 Curtis JR, Jain A, Askling J, *et al*. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. *Semin Arthritis Rheum* 2010; **40**: 2–14 e1.

101 Verstappen SM, Askling J, Berglind N, *et al*. Methodological challenges when comparing demographic and clinical characteristics of international observational registries. *Arthritis Care Res (Hoboken)* 2015; **67**: 1637–45.

102 PCORnet. PCORnet: the National Patient-Centered Clinical Research Network. 2018. https://pcornet.org (accessed May 2019).

103 Fleurence RL, Curtis LH, Califf RM, *et al*. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014; **21**: 578–82.

104 Consortium PCP, Daugherty SE, Wahba S, *et al*. Patient-powered research networks: building capacity for conducting patient-centered clinical outcomes research. *J Am Med Inform Assoc* 2014; **21**: 583–6.

105 Yazdany J, Bansback N, Clowse M, *et al*. Rheumatology informatics system for effectiveness: a national informatics-enabled registry for quality improvement. *Arthritis Care Res (Hoboken)* 2016; **68**: 1866–73.

106 Fleurence RL, Beal AC, Sheridan SE, *et al*. Patient-powered research networks aim to improve patient care and health research. *Health Aff (Millwood)* 2014; **33**: 1212–19.

107 Fleurence RL, Forsythe LP, Lauer M, *et al*. Engaging patients and stakeholders in research proposal review: the patient-centered outcomes research institute. *Ann Intern Med* 2014; **161**: 122–30.

108 Health Measures. PROMIS. 2019. http://www.nihpromis.org (accessed May 2019).

109 Curtis JR, Chen L, Bharat A, *et al*. Linkage of a de-identified United States rheumatoid arthritis registry with administrative data to facilitate comparative effectiveness research. *Arthritis Care Res (Hoboken)* 2014; **66**: 1790–8.

110 El Emam K, Samet S, Arbuckle L, *et al*. A secure distributed logistic regression protocol for the detection of rare adverse drug events. *J Am Med Inform Assoc* 2013; **20**: 453–61.

111 Weber SC, Lowe H, Das A, *et al*. A simple heuristic for blindfolded record linkage. *J Am Med Inform Assoc* 2012; **19**: e157–e161.

112 Kum HC, Krishnamurthy A, Machanavajjhala A, *et al*. Privacy preserving interactive record linkage (PPIRL). *J Am Med Inform Assoc* 2014; **21**: 212–20.

113 Curtis JR, Xie F, Chen L, *et al*. Biomarker-related risk for myocardial infarction and serious infections in patients with rheumatoid arthritis: a population-based study. *Ann Rheum Dis* 2018; **77**(3): 386–92.

114 Curtis JR, Chen L, Greenberg JD, *et al*. The clinical status and economic savings associated with remission among patients with rheumatoid arthritis: leveraging linked registry and claims data for synergistic insights. *Pharmacoepidemiol Drug Saf* 2017; **26**: 310–19.

24

Risk Management

Claudia Manzo[1], Emil Cochino[2], Lubna Merchant[1], and Giampiero Mazzaglia[2]

[1] *Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA*
[2] *European Medicines Agency, Amsterdam, The Netherlands*

Risk management is widely used across a variety of settings to identify, quantify, and characterize risks, and to institute measures to mitigate these risks. Such measures can be used to minimize medical errors in healthcare settings; limit financial liability in the business sector; minimize or eliminate work-related or recreational-related injuries in industrial and leisure settings, respectively; reduce transportation-related accidents in the airline, automobile, and railroad industries; and for many other purposes. Because of the wide-ranging scope of risk management endeavors, the methods of assessing risk vary according to the specific setting. Similarly, the measures used to mitigate risk vary across settings, again depending on the specific risk being managed. While specific measures may vary from setting to setting, at their core, these risk mitigation measures involve a structured approach – generally in the form of some combination of policies, procedures, processes, or engineering solutions – designed to reduce or eliminate one or more specific risks.

Regarding the use of medicines, risk management is used to ensure that the potential benefits of a medicine exceed its potential risks, and to minimize those risks throughout the life cycle of the product. As in other fields, risk management of medicines is not new, though it has received increased attention in the past few decades. Current understanding of the risks of medicines is based on the premise that the risk of a medicine derives not only from the inherent properties of the medicine, but also from how the medicine is used in actual clinical practice. Thus, current risk management efforts are geared toward understanding not only the harm that can result from the intrinsic properties of the medicine, but also the harm that can result from inappropriate use of a medicine in a complex medical care system.

In the context of human medicines in the US, the Food and Drug Administration (FDA) has defined risk management as:

> an iterative process of 1) assessing a product's benefit–risk balance, 2) developing and

---

The views expressed in this chapter are those of the authors, and not necessarily those of the US Food and Drug Administration or the European Medicines Agency.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

> implementing tools to minimize its risks while preserving its benefits, 3) evaluating tool effectiveness and reassessing the benefit–risk balance, and 4) making adjustments, as appropriate, to the risk minimization tools to further improve the benefit–risk balance. This four-part process should be continuous throughout a product's lifecycle, with the results of risk assessment informing decisions regarding risk minimization. [1]

In the European Union (EU), the concept of risk management is established in legislation. Article 1 (28b) of Directive 2001/83 EC, as amended, defines a risk management system as "a set of pharmacovigilance activities and interventions designed to identify, characterize, prevent or minimize risks relating to a medicinal product including the assessment of the effectiveness of those interventions" [2]. Thus, in the EU, risk management incorporates:

- the identification or characterization of the safety profile of the medicinal product, with emphasis on important identified and important potential risks and missing information, and also on which safety concerns need to be managed proactively or further studied (the "safety specification");
- the planning of pharmacovigilance activities aimed at characterizing and quantifying clinically relevant risks, and identifying new adverse reactions (the "pharmacovigilance plan"); and
- the planning and implementation of risk minimization measures, including the evaluation of the effectiveness of these activities (the "risk minimization plan").

As a result, in both the US and the EU, risk management measures are iterative processes frequently leading to the generation of similar data needs and conceptually similar risk management tools.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

All medicines have risks. For marketed medicines, at the time of authorization the benefits of the medicine are judged to outweigh the risks, provided that the medicines are used according to the licensed or approved indication. Knowledge of a medicine's benefits and risks is developed prior to approval, and refined after approval when the exposed population, as a result of "real-world" usage, increases and becomes more heterogeneous, with potential diminishing responsiveness to the beneficial effects and increasing likelihood of detecting adverse drug effects. The traditional tools used to manage the risks of prescription medicines have been the prescription status itself (i.e., whether the drug was approved for prescription-only use or whether it could be obtained without a prescription), labeling for healthcare professionals, and the requirement that pharmaceutical manufacturers monitor and report to regulatory authorities adverse events that occur with use of the medicine once it is marketed. In the past few decades, additional steps or minimization strategies have been undertaken to manage more proactively the risks of certain medicinal products. These measures have included increased communication to patients as well as to healthcare professionals, and measures to restrict, in various ways, the usage of certain medicines. This chapter will explore these efforts in more detail.

The Complexities of the Medication Use System

The medication use system is a complex network of stakeholders, including patients, their families, physicians, nurses, pharmacists, other health professionals, healthcare organizations and healthcare facilities (e.g., hospitals, clinics),

manufacturers, and regulatory agencies. Not only does each individual stakeholder have a role in ensuring the safe use of a medicine, the interactions among the various stakeholders are crucial to ensuring the safe use of a medicine. Thus, risk management strategies must consider not only the individuals and groups, but also the entire medication use system. The complexity of the medication use system implies that individual risk management measures must be directed at the appropriate part or parts of the system specific to the risk being managed. The accurate identification of these parts of the system will vary from one drug to the next, will depend on the specific risk, and will depend on how the medicine is used within the healthcare system. In this context, the approach to risk management must span the entire life cycle, be proactive, be scientifically driven, engage all relevant stakeholders, and consider all environments where the medicine will be used (e.g., hospitals, long-term healthcare facilities, physicians' offices, outpatient home care).

Because the risks of medicines can occur at any point in the complex medication use system, managing the risks of medicines requires that the entire medication use system be involved. Involvement of the entire system can pose challenges, and some parts may be harder to involve than others. It is difficult, though perhaps not impossible, to compel each part of the system to do what it must to manage the risk of a medicine. While the involvement of the entire system is a strength of risk management systems, reliance on each part of the system is a limitation.

Sources of Risk from Medical Products

There are several sources of risks from medical products (Figure 24.1). The known risks of a product are based on prior experience or, in some cases, on the pharmacologic or other properties of the medicine (e.g., the dosage or route of administration). In some cases these risks are preventable, while in others they are not. Preventable risks can occur when a product is administered under a condition of use that imparts a risk that would not be present under a different condition of use. For example, if drug A, when used in combination with drug B, results in an unacceptable risk that is not present when

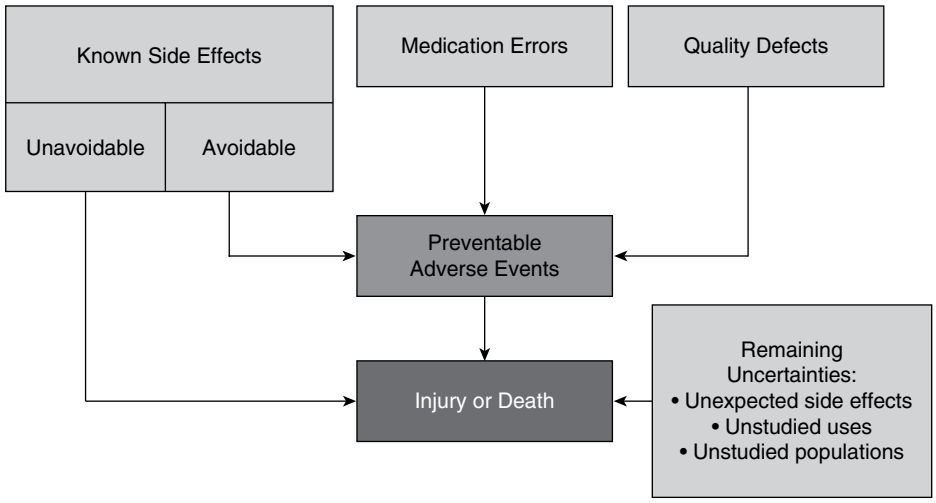


Figure 24.1 Sources of risk from medical products.

either drug is used alone, this unacceptable risk is preventable by ensuring that drug A and drug B are never co-administered. Contraindicating concomitant use is a regulatory step that can be used to warn against concomitant use; actual avoidance of concomitant use can only be achieved by health professionals' adherence to this labeled contraindication. Risk management efforts beyond approved labeling such as additional targeted communication can be used to further minimize preventable adverse events.

Unavoidable risks are those that might occur when all the known necessary conditions for safe use of a product are followed. In these circumstances, risk minimization activities might be directed toward identifying the adverse consequences as early as possible, with the aim of preventing more serious harm. For example, a drug may be known to cause hepatic damage, but its occurrence in a specific patient may not be predictable or preventable. In this case, risk minimization activities might be directed toward regular monitoring of hepatic enzyme levels to identify any hepatic damage as early as possible, and thus to stop or modify the treatment to prevent serious hepatitis or hepatic failure.

In addition, risk management efforts can be used to ensure that medicines are not administered to patients at higher risk for a serious adverse event, or that they are administered only to patients for whom the benefits outweigh the risks, including the unpreventable risks. Thus, removing all risks from the use of all medicines is not the overall goal of managing the risks of medicines. Rather, careful consideration of the benefit/risk balance, both for the individual patient and for the target population, is an important consideration of risk management.

Managing the known risk of medicines is a core activity of risk management programs. For most products, this can be achieved through product labeling; in some cases, as will be discussed later in this chapter, additional steps are needed. Other sources of preventable adverse events are medication errors and, occasionally, injury from product quality defects.

Medication errors (also see Chapter 41) are defined by the National Coordinating Council on Medication Error Reporting and Prevention (NCCMERP) as follows:

> A medication error is any preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of the health care professional, patient, or consumer. Such events may be related to professional practice, health care practice, procedures, and systems, including prescribing; order communication; product labeling, packaging, and nomenclature; compounding; dispensing; distribution; administration; education; monitoring; and use. [3]

A similar definition is in place in the EU [4].

Because they are preventable, medication errors are well suited to risk management efforts. Each year in the US, serious preventable medication errors occur in 3.8 million inpatient admissions and 3.3 million outpatient visits [5,6]. A landmark study published in 2000 estimated that as many as 98 000 people die each year in the US from medical errors occurring in hospitals. The report described medication errors as a significant public health concern that accounts for an estimated 7000 deaths annually in the US [7]. In 2011, the Network for Excellence in Health Innovation reported that outpatient and inpatient preventable medication errors cost approximately $20 billion each year [8]. A recent study in the EU showed a steady increase in the number and proportion of Individual Case Safety Reports (ICSRs) of medication errors in the EudraVigilance database between 2002 and 2015, to a peak of 5% of all ICSRs in the database. Several factors were felt to be responsible for this increase, including greater awareness of the need to report medication errors, guidance on coding for medication

error events, and increased communication to patients and healthcare professionals, as well as the general public, regarding the importance of reporting medication errors [9].

Potential sources of medication errors can include the product's proprietary (brand) name (if it is similar to the name of another medicine, especially if the two medicines have other similar characteristics), the established (generic) name, and the design of the drug product and its container closure system or the container, carton, and packaging. Errors can also occur if the product labeling for healthcare professionals or patients is not clear. Because medication errors can occur anywhere in the medication use system, efforts to minimize the risk of medication error must involve multiple stakeholders.

As already noted, other sources of preventable adverse events include injury from product quality defects. Product quality problems are unusual in both the US and the EU because of the great attention paid to product quality control and quality assurance during manufacturing [10,11]. A discussion of measures to mitigate manufacturing-associated risks is beyond the scope of this chapter.

Because not all the risks of a medicine are known at the time the product is approved, risk management efforts must continue throughout the life cycle of a medicine, as discussed shortly.

Risk Management Strives to Be Scientifically Driven

Risk management plans can be scientifically driven, to the extent that there is available science to inform each component of the plan. The science of risk identification and risk assessment, while still evolving, is well developed, and indeed much of this book describes this science. The science of risk communication is also well developed in general, though its application to communicating the risks of medicines is still being developed (see Chapter 39). The scientific basis of minimizing risks of medicines is much newer, as is the science of assessing the impact of risk management plans. The scientific approach to risk management requires the integration of data from various studies and disciplines that, when taken together, can promote the safe and effective use of a medicine. The scientific approach also compels manufacturers and regulators to examine, throughout the life cycle of the medicine, the critical gaps in knowledge that exist. Such gaps may concern the pharmacologic properties of the medicine, clinical outcomes related to its use, including that in higher-risk populations, or the way the medicine is used in actual practice. Any of these areas could lead to further postapproval studies, the results of which would lead to changes in labeling or other changes that could enhance the safe and effective use of the medicine. However, as noted in the example of cisapride in Chapter 8, changes in labeling do not always result in changes in prescribing practices.

Risk Management Proceeds throughout a Product's Life Cycle

Knowledge about a product's safety profile is always limited to some extent at the time of product approval, because of recognized practical limitations in the drug development process. For example, rare side effects and long-term side effects may not be known when a product is approved because of the relatively small size and short duration of clinical trials. Because some populations are generally not studied in preapproval clinical trials (e.g., pregnant women, children, people with diseases or conditions other than the studied indications for use) or are minimally studied (e.g., older adults), side effects may be discovered if these groups are treated with a product after it goes on the market. Already approved drugs and biologics may receive approval for new uses or in new patient populations, which may necessitate increased vigilance years after the drug was originally

approved. Even after a product has been marketed for a decade or more, uncertainties will remain. For example, a study of new molecular entities approved for use by the US FDA between 2002 and 2014 indicated that safety-related labeling changes were being made as long as 13 years after the products were approved [12]. Because of this life-cycle approach, all stakeholders – patients, practitioners, manufacturers, and regulators – must remain vigilant about the benefit/risk profile of a medicine. Such vigilance is critical for informed decision making, which is an important component of the safe and effective use of medicinal products.

Risk Management Applies to All Medicines

As has already been stressed, all medicines have risks. No medicine is free from harm in all persons who take it under all actual conditions of use. The magnitude, frequency, and severity of risks vary from medicine to medicine. For example, at one end of the spectrum, neutropenia is commonly associated with chemotherapy and other immunosuppressive agents and is a major risk factor for the development of infections. Strategies to monitor, prevent, and/or manage these infections can lead to improved outcomes and will ensure the benefits of these drugs continue to outweigh the risk. At the other end of the spectrum, many topical over-the-counter (OTC) medicines have very few side effects. The management of these risks is clearly much less intense. In the middle of this spectrum are the vast majority of medicines, mainly prescription medicines, for which a measured approach to risk management must be taken.

For most prescription medicines, the most common side effects are generally not life threatening. Rather, many are mild and self-limited. Others are bothersome, and some are so clinically significant that they require the medicine to be discontinued. Examples of these types of side effects whose significance depends upon severity are nausea, headache, and rash. For many medicines, the most serious side effects are relatively rare. Examples of rare, life-threatening side effects are acute liver failure, aplastic anemia, torsade de pointes, progressive multifocal leukoencephalopathy, cytokine release syndrome, and certain serious skin reactions, such as Stevens–Johnson syndrome. Along this continuum are other side effects that may be severe but generally not life threatening, and that are also more common than the most serious side effects such as tendon rupture, narcolepsy, and vision loss.

Over-the-counter medicines are drugs that have been found to be safe and appropriate for use without the supervision of a healthcare professional such as a physician, and they can be purchased by consumers without a prescription. Most OTC medicines are for symptomatic relief of conditions that consumers can diagnose and manage themselves. When these medicines are taken properly, most of their side effects are generally mild. However, there can be serious, even life-threatening or fatal, side effects of OTC medicines when they are not taken properly. For example, acetaminophen (paracetamol), one of the most widely used OTC analgesics, is a generally very safe when taken as recommended on the product's label. Overdose, however, can result in acute severe liver injury, which can lead to acute liver failure, and sometimes death or the need for liver transplantation. While this is a rare complication relative to the widespread use of acetaminophen, the fact that the use is so widespread means that this drug is the leading cause of drug-induced acute liver failure in the US [13].

While much attention is paid to medicines that are known to have life-threatening, fatal, or disabling side effects at therapeutic doses, there are also risks from medicines that do not have these serious side effects when taken properly, but which can cause serious side effects when taken improperly. For example, bromfenac

sodium capsules, an oral nonsteroidal anti-inflammatory agent, was introduced in the US in July 1997 for treatment of pain for 10 days or less. Despite the labeled recommendation for a treatment duration of 10 days or less, many patients received treatment courses of 30 days or longer. The FDA received several reports of hepatotoxicity resulting in death or liver transplantation attributable to bromfenac; in all cases, the patients had taken the medicine for longer than the recommended 10-day duration of treatment. In July 1998, the manufacturer voluntarily withdrew bromfenac sodium capsules from the US market [14]. This example illustrates that improper use of a medicine – in this case treatment durations that exceeded the labeled duration of use – can give rise to serious adverse events.

Risk Management Is a Proactive Process

Risk management systems must be proactive to be optimally effective. The current framework of risk management systems allows a proactive approach in many ways. The ability to identify risks in the preapproval period enables manufacturers to work with regulators on risk management planning and risk minimization strategies during the drug development phase. A proactive approach in the postapproval phase demands that manufacturers, regulators, and practitioners agree on a system to identify new risks, manage known risks, assess the effectiveness of the risk management efforts, and modify them as needed. Traditional pharmacovigilance systems based on spontaneous reports are sometimes referred to as "passive" systems (also see Chapter 10). While such systems can be used in reactive ways, these systems, along with other sources of postapproval drug safety data, can be used in proactive ways to learn as much about the safety of a medicine in as efficient a manner as possible.

Enhanced surveillance systems can be used to proactively monitor the safety profile of a medicinal product following licensure. An example of an enhanced surveillance system in EU is based on the European Medicines Agency's (EMA) seasonal influenza vaccines requirements. Flu vaccine manufacturers are operating surveillance systems enhancing the reporting of adverse reactions by providing vaccination cards to all vaccinees in a region, and raising the awareness of healthcare professionals of the events that might appear following vaccination. By collecting in near real time the usage of the vaccine, the manufacturers are able to monitor the safety profile of the vaccine and rapidly report any emerging signals to regulators, allowing suitable actions to be taken in the same vaccination season. Designs of such a system vary by manufacturer, but range from passive reporting systems with near-real-time estimation of the denominator, to solicited reports monitoring, to observational cohort studies and even clinical trial designs. A carefully designed risk management plan can identify or further characterize risks, communicate and manage risks using evidence-based tools when possible, and assess the effectiveness of these efforts in a proactive way. Like the lifecycle approach noted earlier, the proactive nature of risk management planning demands the constant vigilance of all stakeholders.

Risk Management Activities

Managing the risks of medicines is not a single activity or the province of a single profession or stakeholder group. Rather, it is an iterative process that involves a set of interrelated activities. In broad categories, these activities include risk assessment, risk minimization, and evaluation of risk minimization strategies with adjustments, as appropriate, to the risk minimization strategies to optimize the benefit/risk balance of the medicinal product. These activities occur throughout the product's life cycle, and are adjusted and refined as new risk assessments

provide new information, and as evaluations of risk mitigation activities provide data upon which risk mitigation activities can be improved or modified. Conceptually, for each risk the iterative process starts during clinical development and repeats following authorization.

In the EU, most types of activities are concentrated around important milestones in the lifecycle of the product:

- At initial approval, when the risk assessed using the information gained in the preclinical and clinical phase is balanced with choosing the optimal indication, supplemented with risk minimization recommendations in the product information or additional activities.
- At two other important fixed milestones in the postauthorization phase: (i) when the benefit/risk balance of the product is reevaluated with the first renewal of marketing authorization (MA); and afterwards, (ii) during the first periodic safety update report (PSUR) evaluation (i.e., at 8 years in the life cycle of the product). The important reevaluations of risks are concurrent with the evaluation of the effectiveness of the ongoing risk minimization activities; as a result, recommendations to optimize the safe and effective use of the medicinal product are issued.
- At other minor milestones for risk minimization review and periodic benefit/risk evaluations, or at major changes in product use (e.g., an extension of indication to a new population).

Risk Assessment

Risk assessment consists of identifying, characterizing, and quantifying the risks associated with the use of a medicine, and evaluating their importance in relation to the benefit/risk balance. The nature, frequency, and severity of the risks are assessed. In addition, if possible, the conditions under which the risk is more likely to occur are identified. For example, if a drug causes a serious adverse reaction only when used in conjunction with another specific medicine, it is important to identify this drug–drug interaction, so that risk management efforts can be directed at minimizing the use of the two medicines together.

Risk assessment occurs throughout the premarketing and postmarketing phases of a product's life cycle. Premarket, or preapproval, risk assessment is generally a very extensive process that involves preclinical safety assessments (e.g., animal toxicology testing), clinical pharmacology assessments, and clinical trials. Animal toxicology studies are performed prior to the first human exposure to a new medicine to establish the general toxicity profile of the drug and to guide initial human dosing. Further animal studies continue throughout the drug development process, and address areas such as toxicity (e.g., genotoxicity, carcinogenicity, immunotoxicity, and reproductive toxicity) or safety pharmacology (e.g., cardiovascular system, including electrocardiographic QT interval prolongation, nervous system). Additional animal studies may be needed in specific situations. In addition to animal studies, preclinical testing typically involves the use of *in vitro* bacterial and cell preparations, which can look at effects on enzymes, metabolic pathways, receptors, mutability, and some interactions.

While preclinical research answers basic questions about the safety of a medicine, it is not a substitute for clinical studies to assess how the medicine interacts with the human body. Clinical pharmacologic studies establish the pharmacokinetic profile of the medicine and exposure–response relationships, and can be used to assess drug–drug interactions. Pharmacokinetic characteristics of the medicine under certain clinical situations, such as impaired renal function or impaired hepatic function, can also be assessed. Because proper dosing of a medicine is an important component of the safe use of the medicine, clinical

pharmacologic studies are an important component of a medicine's risk assessment.

Preapproval clinical trials provide the efficacy and safety information that form the basis for an approval decision. The preapproval safety assessment generally quantifies and characterizes the common adverse events associated with a medicinal product. Depending on the number of subjects exposed prior to approval, less common adverse events might also be detected. It is important to pay careful attention to the design of the preapproval safety program to maximize the information gained from clinical trials. The extent of safety information collected prior to approval is a function of the number of patients studied, the duration of treatment, the number of scheduled visits at which safety information is collected, and the specific safety evaluations performed. The design of the preapproval safety data collection effort depends, in turn, on a number of factors, including the novelty of the product, the relative safety of any available alternative treatments, the intended population, the condition being treated, and the intended duration of use. The preapproval clinical safety program should also explore safety-related dose effects and, for chronically administered medicines, the temporal profile of adverse events. It should use the available data to explore unanticipated drug–drug interactions, drug–demographic interactions, drug–disease interactions, and drug–herbal interactions. In some drug development programs, comparative safety data can be obtained if an active comparator is used; however, in the EU and the US the licensing decision relies on the individual benefit/risk profile of the medicine.

Because even large clinical development programs cannot identify all the risks associated with a product, it is imperative that risk assessment continue in the postapproval period, when large numbers of persons will be exposed to the medicine, including many with co-morbid conditions or on concomitant medicines not present in clinical trials. Risk assessment is also imperative for products intended to treat rare diseases where smaller numbers of subjects are studied, or products intended to respond to emergencies, including terrorist events, where approval may be based on evidence of effectiveness from appropriate animal studies when human efficacy studies are not ethical nor feasible. In either scenario, greater risks may be accepted for a treatment that is an advantage over available therapy or addresses an unmet medical need.

Postapproval risk assessment can be based on either nonexperimental data or on clinical trial data. Nonexperimental data include individual case reports of suspected adverse drug reactions (spontaneous reports), case series of such reports, databases of spontaneous reports, disease-based registries, drug-based registries, electronic medical records systems, administrative claims databases, drug utilization databases, poison control center databases, and other public health databases that track usage of the medicine. The use of many of these data sources, and the methods underlying their use, are covered in other chapters of this book (see Chapters 10–17), and will not be considered further here. For the purposes of this chapter, it is important to note that new risks of a medicine will continue to be recognized after the drug is on the market. Some of these risks will be sufficiently serious to alter the benefit/risk balance of the medicine, such that postapproval regulatory action will be needed. Possible regulatory actions include updates to the professional labeling, development of or updates to the patient labeling, use of additional means of communicating risk to patients or healthcare professionals, introduction of additional risk minimization activities (e.g., checklists or monitoring requirements), restrictions on the use of the medicine, or, rarely, suspension or withdrawal of the marketing authorization.

Risk assessment of medicines, in both the preapproval and postapproval phases, often concentrates on the identification of adverse

reactions that are related to the medicine when used according to its labeled instructions. These newly identified adverse reactions can either be an exaggeration of the pharmacologic effect of the drug or an idiosyncratic reaction, the result of a previously unknown drug–drug interaction, or an adverse effect in a specific patient population.

It is also important for risk assessments to identify medication errors and the potential for medication errors (see also Chapter 41) throughout the product's life cycle. The identification and assessment of medication errors are different in some ways from the identification of adverse drug reactions. Proactive risk assessments that reflect human and environmental factors in drug product use should be employed from the earliest stages of product design to help anticipate potential medication errors. Ideally, proactive risk assessments should employ analytic approaches, for example failure mode and effects analysis (FMEA), formative evaluations, or simulated use testing. Considering the end users' needs, environments of use, and contexts of use in the development and design of a drug product alongside commercialization aspects can help reduce postapproval safety issues [15]. After approval, the identification of a medication error generally requires that someone report that an error has occurred, usually in the context of reporting an adverse reaction following the use of the medicine, though the initial report may not elucidate the reason for the error.

Because the medication use system is complex, the mere identification of an error (e.g., the patient received twice the intended dose) is usually not sufficient to understand the reason for the error. Since medication errors are, by definition, preventable events, risk assessment activities must focus on identifying the specific reason(s) for, or cause(s) of, the event. In this situation, it may be essential to conduct a root cause analysis (RCA) to understand the causes (i.e., the how and why) of the problem or medication error. This is an important tool to evaluate postmarketing problems or medication errors, and when evaluating proposed remedies for those problems or errors. Knowledge gained from evaluating the RCA of a known postmarketing medication error can also be applied to the premarket safety assessments of other products. The identified reasons and causes may relate to certain characteristics of the medicinal product itself, to the larger medication use system for the product, or to an interaction of the two. Understanding how and why medication errors occur and what would be the impact on patients are essential to any risk assessment. Only once the specific set of reasons and causes that led to the error are understood can appropriate risk mitigation and risk communication activities be developed.

An EU example of giving the medication to the wrong patient was observed during the clinical development of an autologous advance therapy medicinal product. The root cause identified was a weak identification system, relying on the initials of the patient and treating physician. Once the identification system was improved by using a unique traceability code, from harvest to implantation, no further such medication errors occurred. Additional observed examples of medication errors and the risk minimization activities put in place are provided in the EU good pharmacovigilance practice (GVP) guidelines [16].

Risk Minimization

Risk minimization or mitigation refers to a set of measures or interventions intended to prevent or reduce the occurrence of adverse reactions associated with exposure to a medicine, or to reduce the severity or impact on the patient should adverse reactions occur. Appropriate planning of risk minimization/mitigation allows for medicines with considerable risk to be approved and maintained on the market with a positive benefit/risk balance. The range of risk

mitigation activities varies from one country or region to the next, but certain common themes emerge.

First, many aspects of the modern drug regulatory system are, in fact, risk mitigation activities. The very fact that a medicine has to be approved is, in many ways, the most fundamental risk mitigation activity, in that it prohibits the marketing of medicines that have not been judged to be safe and effective, thus virtually eliminating the risks of medicines being legally marketed for which there is no demonstrated benefit. The requirement that certain medicines be available only by prescription is another form of risk mitigation. The premise underlying the prescription-only status of a medicine is that some medicines are potentially harmful or the method of their use is not safe without the involvement of an appropriately qualified healthcare professional, whose judgment can be used to ensure that, for a particular patient, the potential benefits outweigh the potential risks.

Risk Communication

Communicating information about the benefits and risk of medicines is central to minimizing the risks of these products. Risk communication is a broad field, and a full discussion is beyond the scope of this chapter (also see Chapter 39). Communication has traditionally been directed toward healthcare professionals, but in recent years increasing attention has been paid to communications directed toward patients and consumers.

The principal form of communication to healthcare professionals in the US is the product's approved professional labeling, which is designed to present to the healthcare professional information needed to prescribe the medicinal product in such a way that the potential benefits outweigh the potential risks. In the EU, this professional information is known as the Summary of Product Characteristics (SmPC).

There are several types of information in the professional label that can mitigate risk. First, the label often contains information on those clinical situations in which the drug should not be used, or should be used only with extreme caution. Second, the label contains information about the known risks of the medicine. If prescribers are aware of these risks, they can judge the individual benefit/risk balance for the patient and decide if the medicine is the best choice available; they can also advise patients on the appropriate symptoms to look for when taking the medicine. Upon hearing of these symptoms from patients, prescribers can recognize a potential adverse drug reaction and take appropriate action, such as changing the dose or stopping the medicine. Third, the label contains information about the conditions of safe use of the medicine, such as the proper dosing (including, when applicable, the dose adjustments needed for renal and hepatic impairment or dose adjustments based on age), drug–drug interactions, drug–disease interactions, use in pregnant or lactating women, and use in other specific clinical situations or special populations.

Additional communications to healthcare professionals come in the form of "Dear Healthcare Professional Letters," "Dear Doctor Letters," or "Direct Healthcare Professional Communication." These letters, typically issued by a medicine's manufacturer, are usually one to a few pages in length, and generally focus on specific, newly identified safety information. The nature of the risk is explained, and a summary of the changes to the product label or SmPC is often included. The letter usually highlights actions that the healthcare professional should take in prescribing and dispensing the medicinal product, as well as other measures that can help ensure the product's safe and appropriate use. Full prescribing information is generally attached, so that prescribers can put the new information into context.

In the EU, these letters are reserved for emerging postmarketing safety issues. The letters can be either requested by a regulatory authority,

with the content agreed as part of the safety review, or voluntarily proposed by the manufacturer. In the latter case, the manufacturer must notify the competent regulatory authorities of its intention to distribute the letter, but more often requests the text to be approved, as this enforces the message to healthcare professionals that the text has had regulatory review and endorsement.

Labeling directed toward patients and consumers is also a risk mitigation tool, in that it highlights basic information necessary for the safe use of the product, and often provides instructions for actions to take when certain symptoms are present. Information for patients and consumers is relevant for both prescription and nonprescription medicines.

In the US, information to patients can come in a variety of forms. One common form is product-specific information directed toward patients. This can take the form of approved patient labeling, which is developed by the manufacturer and reviewed and approved by the FDA. Examples of approved patient labeling include the Medication Guide, a Patient Package Insert or Instructions for Use. Medication Guides are used when there is a need to communicate certain safety information, or when certain conditions of safe use must be highlighted. By regulation, the FDA requires that Medication Guides be issued with certain prescribed drugs and biologic products when the Agency determines that:

- the drug product is one for which patient labeling could help prevent serious adverse effects;
- the drug product is one that has serious risk(s) (relative to benefits) of which patients should be made aware because information concerning the risk(s) could affect patients' decision to use, or to continue to use, the product;
- the drug product is important to health and patient adherence to directions for use is crucial to the drug's effectiveness.

The format and content of a Medication Guide as well as the distribution requirements are set forth in regulation [17].

Patient Package Inserts are another form of FDA-approved patient labeling. They differ from Medication Guides in several important respects: (i) their use cannot be mandated, except in certain circumstances; (ii) there are no specified requirements for content and format; and (iii) there is no requirement that they be distributed.

The Instructions for Use or IFU is a form of prescription drug labeling developed for use by patients for drug products with complicated or detailed patient-use instructions. The primary purpose of an IFU is to provide detailed, action-oriented, step-by-step, written and visual instructions in a patient-friendly manner. The IFU is intended to help administer drug products safely and effectively. It is developed by the manufacturer, reviewed and approved by the FDA, and provided to patients when the drug product is dispensed.

Consumer Medication Information is an alternative form of drug-specific patient information used in the US. Unlike FDA-approved patient labeling, Consumer Medication Information is neither developed by the product's manufacturer nor is it regulated by the FDA. Rather, it is developed by independent, commercial, third-party vendors who often sell this and other products to pharmacies, and is then distributed to patients in pharmacies.

In the EU, all medicines are required to have a package leaflet, sometimes referred to as a patient information leaflet, which must be provided to the patient as part of the product packaging, or in exceptional circumstances (e.g., a pandemic) as a separate leaflet, or even as online information. This leaflet is based upon the information provided in the SmPC, but written in patient-friendly language. There is a requirement that the package leaflet reflects the results of readability testing with an appropriate target group of patients/consumers and that the

results be provided to the competent regulatory authorities prior to authorization.

Both the SmPC and the package leaflet are approved by the competent regulatory authority at the time of authorization and require a specific procedure, involving evaluation of the reasons for change and agreement of the authority to the revised text, in order to change the content. Who the competent regulatory authority is depends upon how the medicine is authorized. In the centralized procedure it is the European Commission, whereas for other methods of authorization it is the appropriate regulatory authority in the Member States.

For most prescription medicines, the prescription status, the product information for healthcare professionals and patients, and the labelling are sufficient risk minimization measures to ensure that the benefits of a medicine outweigh its risks. In some cases, additional communication measures are needed.

Regulatory guidelines do not provide clear criteria or recognized algorithms to establish when additional risk minimization is warranted and what tools are best suited to do so. However, accurate recognition of important risks that need to be minimized is the basic starting point, and prioritization of safety concerns should take into account frequency, seriousness, severity, impact on public health and preventability. In addition, the burden of imposing additional risk minimization on stakeholders and the healthcare system should be balanced with the expected reduction of the frequency and/or severity of the targeted risks.

If additional measures are utilized, they are generally designed to address one, or at most a few, specific important risks associated with a medicine. While specific measures may differ from one country or region to the next, these measures may include focused risk information targeted at practitioners that are likely to prescribe the medicine or care for patients who are treated with the medicine. Specific risk information may also be targeted at healthcare professional societies to share with their members. The types of communication can include letters or other educational tools, for example a prescriber checklist or an educational brochure. Communication for patients can include a dosing card for medicines with complicated dosing instructions or a patient alert card. In some cases, the communication is focused principally on the nature of a serious risk, so that patients and prescribers can make an informed decision as to whether the potential benefits outweigh the potential risks in their individual situation. In other cases, the communication focuses on both the nature of the risk (e.g., risk factors, preventability, early signs and symptoms) and the specific steps that can be taken to prevent the event, or particular monitoring that should be carried out to detect the adverse event.

The extent to which specific information about the risk needs to be communicated will depend upon the context in which the product is used. For example, specific activities beyond the product information for a drug which has the potential to prolong the electrocardiographic QT interval will probably not be necessary if the medicine is one intended for use only by cardiologists. However, if the same risk occurs in a drug used by oncologists, additional activities targeted at informing prescribers of the need for periodic monitoring of heart function, and the risk of concomitant use with other QT interval-prolonging drugs, might be appropriate. In certain cases, these additional communication measures may be required as part of a formal risk minimization plan in the EU, or risk evaluation and mitigation strategy (REMS) in the US.

Regulatory agencies have also been engaging in increasing efforts to communicate the risks of medicines. The FDA's primary tool for communicating important new and emerging safety information about a medicine is through a Drug Safety Communication or DSC. DSCs highlight new safety issues that pose potentially serious or life-threatening risks or adverse events, about which the FDA thinks patients and healthcare

professionals should be informed. In establishing this system, the FDA recognized that there is a tension between providing early notification about potentially important safety information on the one hand, and being certain about the findings on the other. Because analyses and interpretations of drug safety information are often not clear cut, communication of findings requires a balanced and impartial approach.

In the EU, the competent regulatory authorities communicate drug safety information using different methods, which depend upon what has been established for the individual country. EMA plays a central role in coordinating these communications. When a drug safety issue has been discussed and agreed by the Pharmacovigilance Risk Assessment Committee (PRAC) or together with the Committee for Human Medicinal Products (CHMP), EMA will issue a public statement on its website giving information about the medicine, the risks, and what is being done to mitigate them. In addition, if a Member State or the European Commission has referred a public health issue to the PRAC or CHMP for a scientific opinion, then the start and reasons for the referral are also announced, as well as the final conclusions.

EMA also coordinates the release of information in Member States to avoid different information appearing at different times across Europe. Information about every medicine (including the risks and benefits) authorized via the centralized procedure is provided online by the European Public Assessment Report (EPAR). This is a scientific document which summarizes the information that EMA evaluated in giving its opinion about the medicine and how the decision to give a positive or negative opinion regarding authorization was reached. The EPAR is updated throughout the life cycle of the medicine to reflect major changes in the product's marketing authorization. In addition, the SmPC, package leaflet, labeling, risk management plan (RMP) summary, and conditions for marketing authorization also appear on the EMA website and similarly are updated throughout the life cycle of the medicine. The requirement for an assessment report to be publicly available also applies to medicines authorized by other routes (e.g., national, mutual recognition, and decentralized procedures), but the exact format in which it appears will vary.

In addition to the mechanisms already listed, manufacturers use a variety of other communication tools to reach practitioners and patients. These can include print and broadcast advertising, websites, and other communications via electronic social media. In many cases, these communications are simply a part of a manufacturer's marketing program; in some cases, however, they may be a formal risk minimization strategy. In the EU, advertising of prescription-only medicines directly to patients is prohibited.

Additional Risk Minimization Strategies

A variety of other minimization strategies can be employed when product labeling and alternative forms of risk communication are not sufficient. For example, drug product design features that predispose end users to medication errors may not always be overcome by product labeling or healthcare provider or patient education. It is therefore preferable to eliminate, or minimize to the extent possible, these hazards from the product design. It is not possible to predict all medication errors; however, medication errors can be minimized by conducting premarketing risk assessments to evaluate how users will interact with the medicinal product in various environments of use within the medication use system. Because medication errors can occur anywhere in the medication use system, efforts to minimize the risk of medication error must involve multiple stakeholders.

In the US, the FDA reviews proposed proprietary names of medicinal products to ensure that these names are not similar in spelling or pronunciation to the proprietary or established names of other medicines. In addition, the FDA

reviews the proposed container labels, carton labeling, packaging, and product design to ensure that these do not have features that could cause or contribute to medication errors.

Similarly, in the EU medicines authorized through the centralized procedure have their invented (or brand) name approved by the (invented) name review group, which checks that there are no products licensed with similar names in the EU which could lead to confusion. In the centralized procedure, the European Commission (based on a scientific assessment by CHMP) issues one license which applies to all the Member States, Norway, Iceland, and Liechtenstein. The layout, format, and wording on the immediate and outer packaging of the product are also reviewed as part of the evaluation procedure of the medicine, and these form part of the authorization. Any change to any of these aspects requires regulatory review and agreement.

There are a variety of other strategies that can be used to mitigate risks associated with a medicine. For healthcare providers, this might include specialized training for certain products, such as one that requires unique administration or insertion (e.g., implants). It may also include materials that facilitate discussions between healthcare providers (HCPs) and patients (or caregivers), such as prescriber–patient agreements. For teratogenic medications, it may include strategies to prevent fetal exposure to a medicine, including required pregnancy testing prior to prescribing or dispensing the medicine to individuals who could become pregnant, as well as contraceptive counseling and assurance that the patient is using measures to prevent pregnancy while taking the teratogenic medicine. Some medicines may require patient monitoring or periods of observation by a healthcare professional after administration, or the medicine may have to be administered in a certain type of healthcare setting that is equipped to manage the serious adverse event.

Pharmacists may play an important role in verifying that the medicine is appropriate for the patient, verifying that certain safe use conditions have been met prior to dispensing the medicine, or providing special instructions for the preparation or administration of the medicinal product. Pharmacists are also well placed to counsel patients at the point of dispensing on the use of their medicines, including other medicines or foods to avoid, special dosing instructions, or storage instructions if, for example, accidental overdose or diversion is a concern.

When determining the specific risk mitigation strategy, it is important to consider its goals and objectives. Interventions should be selected that can achieve those goals and objectives. As a starting point, the risk mitigation interventions undertaken in clinical trials should be evaluated to determine if those strategies were effective. There should be consideration of the healthcare providers likely to prescribe the medicine, the probable setting in which the medicine will be used, the population likely to use the medicine, and if those interventions would be reasonably reproduced in the setting of real-world use. In some cases, such an evaluation may lead to the conclusion that the product can be safely and appropriately used with labeling alone. In other cases, the evaluation may lead to the conclusion that labeling is insufficient, and that additional risk mitigation strategies are necessary to ensure a positive benefit/risk balance.

Some risk minimization strategies may require that prescribing and/or dispensing be limited to those willing to undertake additional steps to ensure safe use of the medicine. Measures that restrict the way a medicine is prescribed or used may introduce complexity and may require coordination with patients, pharmacists, and other stakeholders, particularly when patients move between various care settings (e.g., from an outpatient setting to an inpatient care setting).

In the EU and the US, a controlled distribution system may ensure that the medicinal product is

traceable to the pharmacy dispensing the product, to prevent the misuse and abuse of the medicine. When required, a set of additional risk minimization activities in the form of a controlled access program seeks to minimize an important risk with significant public health or individual patient impact for a product with clearly demonstrated benefits, but which would not otherwise be available. Patients' access to the medicine in such a system is contingent on fulfilling strict requirements before the medicinal product is used, such as clinical testing, completion of an educational program, enrollment in a systematic patient follow-up and data collection system, or availability of the medicine only from approved or certified pharmacies. The requirements can continue during treatment, by imposing specific tests or monitoring procedures. Since a controlled access program has large implications for all stakeholders, the use of these types of programs is generally limited and is guided by a clear therapeutic need for the product based on its demonstrated benefit (e.g., the medicine treats a serious disease without alternative therapies, or it treats patients who have failed existing therapies), the nature of the associated risk (e.g., the risk is life threatening), and the likelihood that this risk can be managed by such a program.

Evaluation of Risk Minimization and Mitigation Measures

Evaluation of risk minimization and mitigation activities is a critical component of a risk management system that follows the risk assessment and risk minimization steps. It is also a relatively new endeavor in the context of the medication use system. The evaluation of a risk mitigation activity aims to ensure that the objectives of the risk mitigation measures are fulfilled and that the activities in place are proportionate, taking into account the benefit/risk profile of the medicinal product and the efforts required to implement these measures. Such an evaluation is closely related to risk assessment activities, but it also differs in the way in which it enables modifications of the initial measures, if warranted, to improve the risk minimization strategy in the context of an iterative process of evaluation, correction, and reevaluation throughout the life cycle of a medicinal product.

In the EU, the pharmacovigilance legislation explicitly requires the active monitoring of the outcome of risk minimization activities contained in the risk management plan, placing an obligation on the manufacturers and regulatory authorities to undertake this activity. In the US, the metrics of a REMS assessment plan are approved in advance of REMS implementation. Assessment reports are submitted by the manufacturers 18 months and 3 years after the REMS is initially approved, and in the seventh year after the REMS is initially approved, with additional dates if more frequent assessments are necessary to ensure that the benefits of the drug continue to outweigh the risks.

There are a number of different conceptual models published as guides for developing efficient methods for measuring the effectiveness of risk minimization strategies. Despite some difference in the approaches, all models have the following common principles:

- Robust risk minimization evaluation is longitudinal in nature.
- A multifaceted assessment is needed for a comprehensive risk minimization evaluation.
- There are some key elements aimed at evaluating the implementation of risk minimization, such as:
  1) enablers and barriers for optimal program delivery and success;
  2) stakeholders' knowledge, attitudes, and perception of risk;
  3) intended and observed clinical behavior.
- Safety outcome data define the ultimate success of a risk minimization program.
- The unintended consequences of risk minimization measures should be taken into account.

First, evaluation of risk mitigation activities can assess if certain processes specified by the risk mitigation strategy are being followed. For example, if the risk mitigation strategy consists of providing patients with specific information about measures that can be undertaken to minimize a particular risk, this first process component of the assessment could consist of determining the proportion of patients who receive the information. The proportion of physicians aware of specific risk mitigation measures is also an example, as well as the proportion of physicians and other prescribers who received and who are using a specific risk minimization tool, such as a dosage card.

The coverage of a risk minimization tool generally does not require properly designed studies on a sampled population, as risk minimization tools are supposed to be distributed to all end users. However, monitoring of tool distribution can be required and estimates of coverage provided on the basis of agreed time frames. Results of inadequate distribution/coverage should lead to reconsideration of the delivery channels employed or help to determine whether a different tool format is required. Tool utilization can be measured using target audience surveys or via proxy indicators such as healthcare professionals' requests for refills of consumable risk minimization items (e.g., checklists and forms). Results can be analyzed as a whole or stratified to show how specific subgroups are performing, with the aim of highlighting areas where coverage, awareness, or usage is poorer and needs to be improved.

Second, the evaluation of risk minimization measures can focus on whether the target audiences understood both the purpose of the risk minimization tools and their key messages. These messages often relate to the safety risks, such as important signs and symptoms, or to actions that should or should not be taken, such as performing laboratory tests or not prescribing a drug to specific subpopulations. Common process indicators to be considered include the proportion of end users correctly responding to specific questions aimed at measuring understanding of the key messages contained in the risk minimization tool. When conducting such studies, scientifically rigorous survey methods should be applied and comprehensive guidelines for research can be retrieved from the published literature.

Overall, the following elements should be considered in the design and implementation of a survey in order to minimize potential biases and optimize the generalizability of the results to the intended population:

- The sampling frame should not be subjected to selection bias (random selection is generally the optimal approach) and recruitment sources should be appropriate (e.g., physicians lists from learned societies, patients lists from registries).
- The design and the administration of the survey questionnaire should ensure a fair and comprehensive evaluation. To build a robust questionnaire, the following three principles should be followed:
  - *Pretesting and validation*: to identify questions that are poorly understood, ambiguous, or evoke undesirable responses, and to avoid leading questions.
  - *Content validity*: the questions should capture all the aspects related to users' comprehension of the risk minimization activity, and should be clear and unambiguous.
  - *Construct validity*: the survey should be able to accurately measure (at different degrees) the risk minimization target audience's knowledge and comprehension.

Another level of the evaluation of risk minimization/mitigation activities can determine whether certain behaviors are being followed. In the previous example, measurements of behavior could assess whether patients, who have read the information they are given, take the specific actions the information recommends. Questionnaire-based stakeholder surveys are

not well suited to assess behavioral modification, because they rely on the respondent's self-reporting and might provoke socially acceptable responses, which have a large impact on the validity of the study's findings. Therefore, this evaluation strategy is better addressed with time-trends analyses, relying mainly on information from electronic medical records (EMRs) included in healthcare databases.

Data from EMRs might in fact provide rapid feedback on the effectiveness of risk minimization strategies, considering that information on drug exposure and patient clinical characteristics (which might define adherence to the recommendations contained in product labeling or educational material) is generally based on secondary data collection in an outpatient setting (i.e., primary care databases).

The ultimate measures of success of a risk minimization program are the safety outcomes. This evaluation can demonstrate whether the introduction of a risk mitigation strategy is correlated with a decrease in frequency of the health outcome that the risk mitigation strategy is designed to minimize. Safety indicators are mostly assessed by estimating incidence rates or cumulative incidence, for instance the number of new adverse reactions in subjects exposed to a certain medicinal product divided by the person-time (or size) of the (exposed) population within a specified time period. Reporting rates (e.g., number of suspected adverse reaction reports attributed to a certain medicinal product over a fixed time period) should only be used with caution, due to the well-known underreporting or due to the lack of patient-level linkage between drug exposure and adverse events. However, measuring reporting rates might be the only way to estimate the frequency of the adverse reaction in the treated population (e.g., with a rare event).

The incidence of adverse events can be evaluated in cohort studies using information from healthcare databases or registries. Disease registries are more suitable for evaluating risk minimization measures, as they may contain information on patients not exposed to the medicinal product, thus potentially providing a background rate of occurrence of the adverse events in the affected population in the absence of exposure to the medicine.

Some aspects of the risk management system, especially those that impose restrictions on the use of a medicine, may be burdensome on the healthcare system and may have unintended consequences. One potential unintended consequence is that the burdens imposed by the system will deter practitioners from prescribing a medicine to patients for whom the benefits outweigh the risks, and for whom that medicine would be the optimal treatment choice. There are few data at this time to address this potential limitation; however, methods to assess burden and ways to reduce the burden might include interviews with stakeholders or the use of focus groups, as well as assessing the workflows associated with implementing risk minimization strategies in various healthcare settings.

It is also important but challenging to identify potential barriers to patient access to the drug related to the implementation of risk minimization strategies. Patient access could be affected if providers choose not to prescribe the drug because they are unwilling to implement the risk minimization strategies. It may be difficult for a patient to find a participating prescriber in their geographic area, thus affecting the patient's access to the drug. Obtaining such data continues to be challenging, because it consists of identifying patients who would be appropriate candidates for a particular medicine, a task that involves clinical judgment, and who did not receive it, a task for which most current pharmacoepidemiologic approaches are not well suited, since they rely heavily on databases of drug exposure.

Assessing the risks of a medicine, instituting risk mitigation measures, and evaluating the impact of those measures form an iterative process. As new risks are identified, new risk mitigation measures may have to be put into effect.

These new measures will then need to be evaluated, and the risk mitigation measures may have to be modified. This iterative process occurs throughout the life cycle of the medicinal product. Pharmaceutical manufacturers are held responsible for the safety of their medicinal products, so it is usually they who fund risk assessments and put in place risk mitigation strategies and evaluations of those strategies. Regulators review the results of manufacturers' testing, proposed risk mitigation strategies, and evaluations of those risk mitigation strategies. In some instances, regulators may conduct independent assessments of drug safety.

As the academic field of pharmacoepidemiology has grown, university-based researchers also conduct drug safety research, either independently of manufacturers and regulators or in collaboration with them. A particular form of collaboration between academic researchers, public health institutions, manufacturers, and sources of public funds that can finance safety research is the public–private partnership. An example in the EU is ADVANCE (Accelerated development of vaccine benefit–risk collaboration in Europe), a publicly (Innovative Medicines Initiative, IMI) and privately (European Federation of Pharmaceutical Industries and Associations, EFPIA) funded project. One of its objectives was to develop conceptual models for public–private interaction, as part of developing best practice and a code of conduct for benefit/risk monitoring of vaccines.

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

Roles of Pharmacoepidemiology in Risk Management

Pharmacoepidemiology can play several roles in risk management. The first, and most fundamental, role is to identify and quantify the risks of a medicine. Identification and quantification of risks can occur using a variety of pharmacoepidemiologic techniques, including clinical trials, spontaneous reports, case series, and observational pharmacoepidemiologic studies.

At the time of approval, clinical trials are the principal source of drug safety data. Clinical trials are well suited to characterizing and quantifying the common adverse effects of a medicine. For most prescription medicines, the majority of common side effects are not so serious that they require risk mitigation measures beyond professional labeling, and/or risk assessment measures beyond routine collection of spontaneous adverse event data.

Though the preapproval testing of a drug is rigorous, and the review of the data is very thorough, there are still some uncertainties about the complete safety profile of a drug when it is brought to market. These uncertainties arise because clinical trials are not well suited to detecting adverse events that are very rare or that occur only after prolonged exposure to a medicine or after a long latency period. In addition, real-world patient populations include patients with a broader range of co-morbidities, on a wider variety of concomitant medicines, and more severe underlying disease than those included in clinical trials. In practice, patients may be treated with dosing regimens, or may receive the medicines for uses that were not studied in clinical trials.

Despite these widely acknowledged limitations of clinical trials with regard to drug safety information, such trials can identify and characterize important drug safety issues that may require specialized risk management efforts. Vigabatrin, an irreversible inhibitor of gamma-amino-butyric acid, was approved in the US in 2009 for the treatment of infantile spasms and for refractory complex partial seizures in adults. It had been already approved in the UK. Several years after that approval, case reports emerged suggesting that vigabatrin was associated with peripheral visual field defects. Subsequent publications

described a slowly progressive, bilateral concentric visual field constriction. Prior to the medicine's approval in the US, the manufacturer conducted, at the request of EMA, a study to characterize better the occurrence of peripheral visual field defects. The primary endpoint was formal visual field testing every 4–6 months for 3 years. This measure allows for the detection of visual field defects that may otherwise not be detected in routine practice. Data from 524 patients who had at least one conclusive visual field measure during the course of the study were analyzed. Among adult patients, 35.6% had at least one occurrence of bilateral concentric peripheral visual field constriction; 24.6% had at least two occurrences. The corresponding figures in children were 20.0% and 15.3%, respectively [18]. These data demonstrate that clinical trials can be used to characterize and quantify specialized drug safety questions. In the US, the product was approved with a REMS to address this serious safety concern, as described later in this chapter.

An important use of pharmacoepidemiology is to measure how medicines are used in practice, especially if that is under conditions that can lead to adverse outcomes. Examples of pharmacoepidemiologic findings that could signal that a product is not being used appropriately include a finding that a medicine is being prescribed concomitantly with a contraindicated medicine, a finding that a drug is being used in a population of patients for whom the potential benefits do not outweigh the potential risks, and a finding that a medicine is frequently prescribed for a duration of treatment that is associated with an increased risk of serious adverse events. There are many other potential scenarios that can be examined. For these analyses, drug utilization databases, electronic medical record systems, and other administrative healthcare data, especially those with longitudinal patient-level data, are often useful.

Drug utilization data and medical records can also be used to identify medication errors. For example, the FDA utilized the Sentinel system to investigate whether name confusion errors could be identified by assessing the presence and absence of on- and off-label indications in claims data. Sentinel is the FDA's postmarket medical product safety surveillance system, which utilizes electronic claims and medical record data [19]. The FDA used one of Sentinel's analytic tools to identify potential prescribing and dispensing errors resulting from confusion between two medicinal products with similar proprietary names, Brilinta® (ticagrelor), an antiplatelet medicine, and Brintellix® (vortioxetine), an antidepressant medicine. The study was conducted by assessing the presence and absence of on- and off-label indications in the claims data for both products. Sentinel was used to identify new users of Brintellix, and separately new users of Brilinta, between September 30, 2013 and September 30, 2015. Members of all ages were included and had to be enrolled with medical and pharmacy coverage for ≥365 days prior to the dispensing date. Brintellix users overall were identified, including those who did not have an on- or off-label indication for Brintellix and had an on- or off-label indication for Brilinta. The reverse was done for Brilinta. The FDA identified 18 793 new users of Brintellix, of which 71 (0.4%) had no on- or off-label indication for Brintellix but had an on- or off-label indication for Brilinta in the 365 days prior to the dispensing. The claims profile review included 17 of these users, of whom 5 had no history of an on- or off-label indication for Brintellix and no dispensing of a drug in the same class as Brintellix, suggesting they may be true medication errors. The FDA identified 19 936 Brilinta users overall, of whom 90 (0.5%) had no on- or off-label indication for Brilinta but did have one for Brintellix; 21 of these were in the claims profile review; 8 may be true medication errors. Thus, this study indicated that a claims-based algorithm can be developed to identify potential name confusion medication errors in Sentinel using a combination of routine tools and claims profile review [20].

Examination of the prescribing patterns of long-acting beta agonists has been informative in the management of the risks of these agents. Long-acting beta agonists (LABAs) are indicated to treat, among other things, asthma. However, large clinical trials and meta-analyses of clinical trials have demonstrated that patients treated with these agents have a higher risk of asthma-related death, intubations, and hospitalization. Some data suggest, but do not prove, that this risk is mitigated if the LABAs are used in conjunction with inhaled corticosteroids (ICS) or other asthma controller medications. The National Asthma Education and Prevention Program's (NAEPP) Expert Panel Report 3 (EPR-3) recommends low-dose inhaled corticosteroids (ICS) as the preferred treatment for mild, persistent asthma, and that LABAs be reserved for patients whose asthma is uncontrolled by ICS monotherapy [21]. Friedman and colleagues examined drug-use patterns and clinical indicators of disease severity from a commercial insurance database to characterize use of LABA/ICS combination drugs (nearly all LABA use for asthma is in the form of a LABA/ICS combination product) [22]. Among 87 459 patients with a new prescription for a combination LABA/ICS product, 69.1% had no prior prescription for an ICS and no indicator of disease severity other than mild disease in the 365 days prior to the LABA/ICS prescription. These data suggested that LABAs were not being used in accordance with the national guidelines and that many patients were being exposed unnecessarily to the risks of LABAs [23].

A third application of pharmacoepidemiology is to provide population-based assessments of the causes and contexts in which known harm from medications can occur. For these analyses, one or more public health databases may be especially helpful. These databases can be used to estimate the burden of a given drug-related toxicity in the population. Because they are designed for the public health purposes of quantifying health and harm in society, projected national-level estimates are often available. They are especially useful when considering risk from a class of medications, or from a specific ingredient when it is a component of multiple medications.

As noted earlier in this chapter, acetaminophen is one of the most widely used medicines in the US, available in several single-ingredient and multi-ingredient, OTC, and prescription products. Although acetaminophen is generally safe when used as directed, misuse and overdose can cause acute liver failure, sometimes resulting in liver transplantation or death. Overdoses can be either intentional or unintentional. To provide context for risk mitigation activities, Nourjah and colleagues at the FDA examined several national databases to quantify this problem [24]. Using the National Hospital Ambulatory Medical Care Survey (NHAMCS), they determined that an estimated 56 000 emergency department visits occurred annually between 1993 and 1999 for acetaminophen overdoses; an estimated 12 650 of these overdoes were unintentional. Using data from the National Hospital Discharge Survey (NHDS), they estimated that for the years 1990 to 1999, there were 26 000 hospitalizations annually for acetaminophen overdoses, with 2240 of these related to unintentional overdose annually. Using the National Multiple Cause of Death Files, they estimated that 458 deaths occurred annually from acetaminophen overdose – 100 of which were due to unintentional overdose. These data provide quantitative information on the overall magnitude of acetaminophen overdoses in the US, as well as on the proportion of the overdoses that occur unintentionally. Data such as these are critical, not only for targeting risk mitigation interventions, but also for monitoring their impact once interventions have been implemented. In the UK, regulators have undertaken specific risk mitigation measures related to the potential dangers of acetaminophen, such as labeling recommendations for all medicines containing acetaminophen (paracetamol) and

restricting the amount of paracetamol that can be sold without a prescription to a patient to a maximum of 8 g per sale [25,26]. In the US, a boxed warning was added to prescription drug products containing acetaminophen highlighting the potential for severe liver injury. The FDA also requested all drug manufacturers of oral prescription products to limit the maximum amount of acetaminophen to 325 mg per tablet [27].

A fourth, emerging role of pharmacoepidemiology in the field of risk management is the assessment of risk mitigation efforts. Of all the ways in which pharmacoepidemiology can be used in risk management, understanding the best ways to assess risk mitigation efforts is the least developed. There are many challenges. First, for an effective evaluation, the risk mitigation activity must have a clearly defined goal that is relevant and measurable, even if prespecified criteria for success or failure are not established. Goals that are based on vague or imprecise metrics generally cannot be measured, and even if they are measurable, interpretations of the findings would be difficult.

Second, as already noted, assessing the effectiveness of a risk mitigation strategy can be conducted at several levels, including processes, behaviors, and health outcomes. While the traditional methods of pharmacoepidemiology may be used to assess observed behavior and health outcomes, it is quite likely that additional methods, such as those used in social sciences and health policy and management fields, may be needed for the first two levels (process and behavior).

Third, it is important to understand the relationship between each component of the risk mitigation strategy and the desired health outcomes. It is possible that practitioners and patients adhere to the processes and exhibit the behaviors desired by the risk mitigation strategy, but that the health outcome of interest is not improved or is difficult to measure. Alternatively, it is possible that practitioners and patients do not adhere to the processes or exhibit the desired behaviors, but the desired health outcome (e.g., a reduction in the specific risk) is achieved, perhaps because of other interventions or factors that were not part of the risk mitigation strategy. In either case, a critical examination of the risk mitigation strategy would be necessary.

The analysis of a risk mitigation strategy for isotretinoin illustrates how pharmacoepidemiology can be used to assess the impact of various program measures on program effectiveness. Approved in the US in 1982, isotretinoin is a medicine that is uniquely effective in the treatment of severe, recalcitrant nodular, cystic acne. It can cause severe birth defects and intrauterine fetal deaths and, to minimize the risk of fetal exposure, a series of risk management efforts have been implemented throughout the life cycle of this medicine. At the time of approval, risk messages for patients and prescribers were included in the approved labeling and initial marketing materials. In 1988, the manufacturer implemented the Accutane Pregnancy Prevention Program (PPP). The program included strengthened labeling, targeted education, reminder tools, patient informed consent forms, and patient and prescriber surveys to assess compliance with the program.

By 2000, the FDA had concluded that the PPP was not effective in minimizing exposure during pregnancy. The manufacturer then developed the System to Manage Accutane-Related Teratogenicity (SMART). An essential feature of this program was a "qualification" sticker that was to be attached to a written prescription for isotretinoin, which was to indicate adherence to certain program-mandated steps by prescribers and female patients. Prescribers were to read certain material about the teratogenic effects of isotretinoin and sign a letter attesting to their understanding of the measures to minimize fetal exposure to isotretinoin. Voluntary prescriber education was made available and encouraged.

Upon receipt of this letter by the manufacturer, the prescriber was eligible to receive qualification stickers. Qualification of female

patients involved multiple steps. The first step consisted of education about the teratogenic effects of isotretinoin, signing a consent form indicating understanding of the risks associated with the use of isotretinoin during pregnancy, and documentation of an initial negative serum or urine pregnancy test. In the second step, prescribers counseled sexually active women to select and use simultaneously two forms of effective contraception control, from a list of acceptable primary and secondary methods outlined in the SMART program, for one month prior to initiation of isotretinoin treatment, during treatment, and for one month after discontinuation of treatment. The third step was a confirmatory negative pregnancy test within seven days before the actual start of treatment. When each of these steps had been met, the patient was qualified, and was to present a written prescription with a qualification sticker to the pharmacist, who was to dispense isotretinoin only if the qualification sticker was present. The supply was limited to 30 days, and was to be accompanied by a Medication Guide. Before additional isotretinoin could be dispensed, women were again to qualify by having a negative serum or urine pregnancy test.

Prior to implementing SMART, the manufacturer agreed to evaluate the program's effectiveness during the first year [28]. A Pharmacy Compliance Survey found that, after the third month of the program, compliance with the requirement for a "qualification" sticker was generally high, above 99% for urban pharmacies and above 90% for rural pharmacies. The proportion of stickers that contained information on patient sex and qualification date was similarly high. A voluntary patient survey, which enrolled 21–26% of patients during the first four quarter-years of SMART, revealed that 9% of women reported signing no consent form, 81% indicated they received a Medication Guide, and 90% reported receiving a qualification sticker on their prescription. Among women aged 15–45 of child-bearing potential, 91% reported at least one pregnancy test and 66% reported two pregnancy tests prior to the initiation of treatment.

In further analyses, FDA staff examined the relationship of a qualification sticker to pregnancy testing and use of birth control measures. For the pregnancy test analysis, across 4400 prescriptions, a qualification sticker was present for 4300 and not present for 100. The frequency of pregnancy testing when a sticker was present was 91%; the corresponding frequency when a sticker was not present was 90%. For the birth control analysis, across 1788 prescriptions, a qualification sticker was present for 1715 and not present for 73. The frequency of reported birth control use testing when a sticker was present was 97%; the corresponding frequency when a sticker was not present was 96%.

The qualification sticker in the SMART program was, in some ways, designed to be a surrogate marker for important conditions of safe use of isotretinoin. The analysis shows that despite reasonably high compliance with the placement of a qualification sticker on a prescription, this measure yielded information no different from the lack of a sticker with regard to two important conditions of safe use: the undergoing of pretreatment pregnancy tests and the use of acceptable methods of birth control.

These findings indicate that process measures that are a surrogate for clinical events need to be validated. The results of the evaluation of the SMART program led to the development of iPLEDGE®, a single, shared, restricted distribution program that includes all isotretinoin products. iPLEDGE requires the participation and enrollment of prescribers, patients, pharmacies, and wholesalers. Unlike SMART, iPLEDGE links the dispensing of isotretinoin to the documentation of negative pregnancy tests, prescriber confirmation that contraceptive counseling has occurred, prescriber and patient identification of contraceptive methods chosen, and demonstration of patient comprehension of isotretinoin risk and measures intended to mitigate risk. When all elements have been confirmed,

iPLEDGE provides a "risk management authorization" to the pharmacist. While it is not known if the total number of fetal exposures to isotretinoin has decreased under the iPLEDGE program, there is greater certainty that patient counseling and pregnancy testing are being conducted, and that patients are made aware of the risks and agree to appropriate measures to prevent pregnancy during treatment [29].

Another role for pharmacoepidemiology is in the area of assessing risk communication. The assessment of communication is a broad endeavor, and can involve many disciplines and approaches. A survey of the evidence base for the factors that can contribute to improved content and format of patient-oriented prescription drug labeling identified randomized clinical trials, surveys, questionnaires, interviews, and other methods used to assess readability and understanding, though it noted that little evidence existed linking label design or content to measurable health outcomes [30].

To assess the relationship of various communication strategies to health outcomes, Shrank and colleagues took advantage of a deliberate effort by one large pharmacy chain to improve its patient labeling [31]. They then used administrative claims data from one insurance carrier in two states in the US to look at various health outcomes for patients to whom outpatient medicines for one of nine chronic conditions were dispensed. Because these data contained detailed information on the specific pharmacy at which the medicines were dispensed, they could examine the impact of the new labeling strategy, which was linked to one specific pharmacy chain, on health outcomes. Health outcomes of interest included outpatient, emergency department, and inpatient health services use. The sample included 23 745 users of the pharmacy which introduced the newly designed labeling, and 162 369 matched patients who used other pharmacies.

The study found that the introduction of the modified labeling was not associated with a significant change in the rates of outpatient health services use (event rate ratio: 0.53, 95% confidence interval [CI]: 0.15–1.86) or inpatient and emergency department care (event rate ratio: 0.88, 95% CI: 0.62–1.24) among users of pharmacies that incorporated the modified labeling compared to users of pharmacies that did not incorporate the new labeling. However, the 95% confidence intervals include clinically important event ratios, which suggest that insufficient power, and not failure of the intervention, may account for the lack of a statistically significant finding. The authors noted the challenges in developing health literacy interventions that can have a measurable impact on health outcomes.

Currently Available Solutions

Regulatory Framework in the US

It is important to distinguish the broad strategies used to manage the risks of medicines from the specific legislative initiatives that are often associated with risk management. Specifically, the latter are generally a subset of the former.

The Food and Drug Administration Amendments Act (FDAAA) of 2007 created section 505-1 of the Federal Food, Drug, and Cosmetic Act (FDCA), which authorizes the FDA to require a risk evaluation and mitigation strategy or REMS for certain drugs if it determines that a REMS is necessary to ensure that the benefits of the drug outweigh its risks [32]. A REMS is a required risk management plan that utilizes tools beyond routine labeling to ensure that the benefits of a drug outweigh its risks. The FDA can require manufacturers to develop and comply with the REMS if specific statutory criteria are met. These provisions became effective in 2008. The REMS authorities apply to prescription products approved under New Drug Applications (NDAs) and Abbreviated New Drug Applications (ANDAs), as well as products approved under Biologics License Applications (BLAs).

Prior to the initial approval of an application, the FDA offices responsible for review of the

medicinal product and for postapproval safety review determine whether a REMS is needed to ensure that the benefits of the drug outweigh its risks. In the postapproval phase, the FDA may determine that a REMS is needed if it becomes aware of new safety information after the drug was approved, and determines that a REMS is necessary to ensure that the benefits of the drug outweigh its risks. New safety information may be derived from a clinical trial or study, adverse event reports, published literature, or other scientific data deemed appropriate by the FDA about a serious risk or unexpected serious risk associated with the use of the drug. This may include information based on a new analysis of existing data or an assessment of the effectiveness of an approved REMS. The FDA makes decisions about requiring a REMS as part of a benefit–risk determination for a drug after an evaluation that includes consideration of certain statutory factors (listed in Table 24.1) [33].

All REMS for NDAs and BLAs must include a timetable for assessment of the REMS and may include the following additional elements:

- A Medication Guide (MG) or a patient package insert (PPI)
- A communication plan
- Certain packaging and safe disposal technologies for drugs that pose a serious risk of abuse or overdose
- Elements to assure safe use (ETASU)
- An implementation system

A Medication Guide provides FDA-approved patient-focused labeling, and may be required as part of a REMS to inform patients about serious risks associated with the product; it may also be used to provide patients with information necessary for the safe use of the product. A communication plan consists of FDA-approved materials used to aid the implementation of the REMS and/or inform healthcare providers about serious risks of a product. FDA may require certain packaging or safe disposal systems be made available for opioids and other drugs that pose a serious risk of abuse or overdose if the FDA determines that such packaging may mitigate such risks [34]. Elements to assure safe use (listed in Table 24.2) are required if they are necessary to mitigate a specific serious risk listed in the labeling of a product, thus enabling access for patients to drugs that would otherwise not be approved. The FDA may also require an implementation system for REMS with certain ETASU. An implementation system requires the application holder to take reasonable steps to monitor, evaluate, and improve implementation of ETASU by healthcare providers and other participants.

The minimal timetable for assessment of a REMS includes assessments by 18 months, by 3 years, and in the seventh-year post-REMS

Table 24.1 Factors the US Food and Drug Administration must consider when determining the need for a risk evaluation and mitigation strategy.

Estimated size of the population likely to use the drug involved
Seriousness of the disease or condition that is to be treated with the drug
Expected benefit of the drug with respect to such disease or condition
Expected or actual duration of treatment with the drug
Seriousness of any known or potential adverse events that may be related to the drug and the background incidence of such events in the population likely to use the drug
Whether the drug is a new molecular entity

Source: Federal Food, Drug and Cosmetic Act, section 505-1(a)(1).

Table 24.2 Risk evaluation and mitigation strategies in the US: elements to assure safe use.

A) Healthcare providers who prescribe the drug have particular training or experience or are specially certified
B) Pharmacies, practitioners, or healthcare settings that dispense the drug are specially certified
C) The drug be dispensed to patients only in certain healthcare settings, such as hospitals
D) The drug be dispensed to patients with evidence or other documentation of safe use conditions, such as laboratory test results
E) Each patient using the drug be subject to certain monitoring
F) Each patient using the drug be enrolled in a registry

Source: Federal Food, Drug and Cosmetic Act, section 505-1(f)(3).

approval. The FDA may require more frequent assessments that may be specified in the REMS.

Regulatory Framework in the EU

The requirement for risk management in the EU is specifically included in legislation. The term "medicinal product" is also defined in the legislation and includes both chemical and biologic medicines.

Article 6 of Regulation (EC) No 726/2004 as amended and Article 8 of Directive 2001/83/EC as amended lay down the requirements for the documents to be included in an application for the authorization of a medicinal product for human use. A description of the pharmacovigilance system is developed by the manufacturer for a marketing authorization in the format of a pharmacovigilance system master file (PSMF); the legal requirement is to maintain and make available upon request the PSMF (Directive 2010/84/EU). It is a requirement of the marketing authorization application that summary information about the pharmacovigilance system is submitted to the competent authorities; this summary includes information on the location of the PSMF.

The manufacturer is also responsible for developing and maintaining product-specific risk management systems. The terms "pharmacovigilance systems" and "risk management systems" mentioned in the legislation may cause confusion. A pharmacovigilance system refers to the measures that a company puts in place to meet the requirements in the legislation for pharmacovigilance, designed to monitor the safety of authorized medicinal products and detect any change to their benefit/risk balance. These requirements include the need to have within the company an "appropriately qualified person responsible for pharmacovigilance." This person must reside within the EU, Norway, Iceland, or Liechtenstein, and is known as the qualified person responsible for pharmacovigilance (EU QPPV). The pharmacovigilance activities that they and the company are responsible for can be summarized briefly as:

- Setting up a system to ensure that all reports of suspected adverse reactions are collected, collated, and accessible.
- Preparing and submitting reports to the authorities of both individual adverse reactions and periodic safety update reports, as specified in the legislation and guidance.
- Providing the authorities with any requested or any other information that relates to the benefits or risks of the medicinal product, including drug utilization data.

The means whereby this is achieved is known as the pharmacovigilance system. A pharmacovigilance system is therefore company specific and would include the adverse reaction database, the EU QPPV, and the various processes, standard operating procedures (SOPs), and so on, by which an individual company ensures compliance with pharmacovigilance legislation. The requirements for the description of the pharmacovigilance system are described in Module II – Pharmacovigilance

system master file of the Guideline on good pharmacovigilance practices.

Whereas the pharmacovigilance system applies to the people and processes in a whole company, a risk management system is usually substance specific, or, if the manufacturer has authorizations for more than one product containing the same active substance, the risk management system will be substance specific. It describes the important risks and missing safety information pertaining to a particular product; how they will be investigated and characterized further; how new risks will be identified; and the risk minimization activities which will be put in place to prevent or mitigate them.

All marketing authorization applications are required to include a description of the risk management system in the form of the Risk Management Plan, submitted for assessment. The guidance on the format and content of the EU RMP is provided in the Guideline on good pharmacovigilance practices Module V – Risk management systems, and the Guidance on the format of the risk management plan in the EU.

The aim of an RMP is to document the risk management system considered necessary to identify, characterize, and minimize a medicinal product's important risks. As knowledge regarding a medicinal product's safety profile increases over time, so will the RMP change.

The RMP structure requirements are described in EU legislation, Commission Implementing Regulation (EU) No 520/2012. The three main parts of the RMP mirror the essential activities in risk management:

- Part II – Safety specification, aimed at the identification of safety concerns for the product (i.e., important identified risks, important potential risks, and missing information).
- Part III – Pharmacovigilance plan (including postauthorization safety studies), describing the plan to further characterize the safety concerns of the product.
- Part V – Risk minimization measures (including evaluation of the effectiveness of risk minimization activities), describing the plan to prevent, minimize, and manage the important risks of the product.

The content requirements for the RMP are risk proportionate. For example, the RMP Part II – Safety specification of a medicinal product containing a new active substance will include more information than one of a generic or hybrid product, as the uncertainty is greater at the approval of the first product containing a new active substance. However, the list of safety concerns is determined by the risks associated with the substance or the specifics of each product (e.g., particular formulation). Similarly, the postmarketing requirements for pharmacovigilance and risk minimization activities reflect the safety profile of products, and for common risks equal requirements apply.

Routine Risk Minimization

EU GVP Module V defines routine risk minimization as those activities which apply to every medicinal product and lists all the available tools. These "routine" risk minimization activities are as follows:

- Product information:
  - Summary of Product Characteristics (SmPC), targeted at HCPs, including routine risk communication messages and routine risk minimization activities recommending specific clinical measures to address the risk.
  - Patient Information Leaflet (PIL), targeted at patients/caregivers.
- Labeling (the immediate and outer packaging of the medicine) and pack design [35].
- Pack size, ensuring that patients see an HCP after the use of the doses in the pack.
- Legal status of a medicine (i.e., whether subject to medical prescription or not subject to medical prescription).

Medicines subject to medical prescription may have a further limitation by being categorized as being on special or restricted prescription.

These further categories are defined in Article 71 of Directive 2001/83/EC as amended. Some Member States have within their national legislation the ability to specify further the use of a medicine, but this is not common to all EU countries. For example, in the UK, a medicine that is not subject to medical prescription can be classified as being available only when a pharmacist is present or as suitable for sale without a pharmacist.

A medicine within the "restricted" medicine category will have the details of the restriction in the SmPC, and will be:

- reserved for treatments which can only be followed in a hospital environment, because of its pharmaceutical characteristics or novelty or in the interests of public health;
- used in the treatment of conditions which must be diagnosed in a hospital environment or in institutions with adequate diagnostic facilities, although administration and follow-up may be carried out elsewhere;
- intended for outpatients, but its use may produce very serious adverse reactions, requiring a prescription to be drawn up as required by a specialist and special supervision throughout the treatment.

All these routine risk minimization activities are part of the authorization for a medicine and, for centrally authorized medicines, are contained within the Annexes to the Commission Decision, translated into all EU languages (all translations are at http://ec.europa.eu/health/documents/community-register/html/reg_hum_act.htm?sort=a). They are thus legally binding on the manufacturer.

Additional risk minimization activities

Centrally authorized medicines may have additional risk minimization measures specified as part of the marketing authorization. Article 9 (4)(c) of Regulation (EC) No 726/2004 requires the CHMP to attach to the scientific opinion it gives to the European Commission "details of any recommended conditions or restrictions with regard to the safe and effective use of the medicinal product." These recommendations will be adopted by the European Commission and form part of the legally binding conditions of the Marketing Authorization. In exceptional cases, the European Commission will also adopt a decision related to Article 127 (a) of Directive 2001/83/EC, directed to the Member States, describing the responsibilities of national competent authorities in ensuring that the additional risk minimization measures are implemented in the Member States in accordance with defined key elements.

The legislation refers to recommended conditions or restrictions with regard to the safe and effective use of the medicinal product. This permits any measure deemed necessary by the PRAC and the CHMP (and the European Commission) to be a legally binding condition of the marketing authorization with which the manufacturer must comply.

Each condition will usually state what must be achieved, but how this will be done remains flexible. For example, the conditions of the Marketing Authorization may specify that the manufacturer should set up controlled distribution. Because of the differences in the way healthcare is delivered in each of the Member States, there are at least four different ways in practice in which distribution is controlled across the EU. Specifying the end rather than the means allows for this flexibility.

The conditions may stipulate that certain information is provided to physicians, patients, or both. Typically, they will state what the information must contain, but the format it is presented in, how it is provided, and the particular phrasing of the information are usually left flexible. An exception to this was the risk minimization activities related to 5-aminolevulinic acid hydrochloride. These required that the product only be used by "experienced neurosurgeons conversant with surgery of malignant gliomas and in-depth knowledge of functional

brain anatomy who have completed a training course in fluorescence-guided surgery." The conditions of the Marketing Authorization required the manufacturer, in agreement with the competent authorities in the Member States, to implement training courses prior to launch of the product. The conditions included considerable details on exactly what should be included in the training course, the qualifications and experience needed to become a trainer, and the minimum requirements for a training center. Because the Commission Decision also required the Member States to ensure that these conditions were implemented in their territory, this meant that they had to put in place measures to restrict the use of the product to appropriately trained neurosurgeons. Since not all Member States had centers involved in clinical trials, training centers did not exist initially in all countries. Consequently, training of neurosurgeons from those countries in the particular techniques necessary to use the product safely would need to take place in another Member State. This external training would also need to be continued until sufficient expertise had been developed to enable the specific requirements for trainers and training centers to be met in each country. This case illustrates the fact that very stringent conditions can be set for risk minimization while still allowing flexibility in how it is achieved.

The ability to set conditions and restrictions does not apply to medicines only at the time of authorization. If during the life cycle of a medicine it becomes apparent that additional risk minimization activities are necessary, then these can be made conditions of the Marketing Authorization. In the same way, if it becomes apparent that the risks in real-world usage are not as great as estimated at the time of authorization or when the additional minimization activities are incorporated in routine clinical practice and are well adhered to, it is possible to remove conditions or restrictions. A critical review milestone is the five-year renewal, when an active review of the suitability of risk minimization activities is performed by the manufacturer and PRAC, and changes to the conditions for the safe and effective use of the product are proposed as needed.

Risk Management Example in the US: Alemtuzumab

As mentioned previously in this chapter, the FDA can require a REMS at the initial approval of a medicinal product if it is determined that such a strategy is necessary to ensure that the benefits of a drug outweigh its risks, or postapproval if the FDA becomes aware of new safety information following approval of the product and determines that a REMS is necessary to ensure the benefits outweigh the risks.

Alemtuzumab (with the brand name Lemtrada) is a CD52-directed cytolytic monoclonal antibody indicated for the treatment of patients with relapsing forms of multiple sclerosis (MS). It depletes T and B lymphocytes via binding to the cell surface antigen CD52, and it is believed that any effects in MS are mediated through this action. Alemtuzumab was originally approved (with the brand name Campath) in 2001 for the treatment of patients with B-cell chronic lymphocytic leukemia (B-CLL) without a formal risk management plan. However, the manufacturer removed alemtuzumab (as Campath) from the commercial market prior to its approval as Lemtrada for the treatment of MS.

Alemtuzumab for the treatment of MS has a unique dosing and administration schedule of two treatment courses. The first treatment course is administered via intravenous infusion on five consecutive days, and the second course is administered on three consecutive days, 12 months later. Subsequent treatment courses may be administered, as needed, at least 12 months after the last dose of any prior treatment course [36]. Besides its more immediate effects, infusion-associated reactions (IAR) occurring during the infusion and 24 hours thereafter, alemtuzumab is associated

with longer-term effects that can occur 1–5 years after administration. Secondary autoimmune diseases are the most important adverse events associated with alemtuzumab therapy, predominantly affecting the thyroid, kidneys, and thrombocytic function in approximately 30–40% of patients taking the product [37]. Serious and life-threatening stroke has been reported within 5 days of alemtuzumab administration.

Because of the risks associated with alemtuzumab, its approved US label states that "the use of Lemtrada should generally be reserved for patients who have had an inadequate response to two or more drugs indicated for the treatment of MS." It was also approved in the US with a REMS. The specific goals of the alemtuzumab REMS are provided in Table 24.3. This REMS includes several components to address the short- and long-term risks associated with its use [38].

Only healthcare providers certified in the REMS program are able to prescribe alemtuzumab. Certification includes an educational component that instructs prescribers about the autoimmune risks associated with alemtuzumab, including the risk of idiopathic thrombocytopenia purpura as well as other cytopenias, glomerular nephropathies, thyroid disorders, stroke, and malignancies; how to recognize the clinical presentation of these adverse events; and how to monitor for these events. Prescribers also agree to monitor patients periodically for at least 48 months following the final alemtuzumab infusion.

Alemtuzmab can only be administered by certified healthcare facilities that have on-site access to equipment and personnel who are trained and capable of managing infusion-related reactions, including anaphylaxis and cardiac and respiratory emergencies. Patients are premedicated with high doses of corticosteroids, and possibly antihistamines and/or antipyretics. They are monitored during the entire infusion and for a minimum of two hours following the infusion, or longer if clinically indicated.

Patients who are considered candidates for treatment with alemtuzumab must be enrolled by their provider in the program and counseled about the risks associated with the drug. The initial enrollment form captures baseline patient information, including baseline laboratory data. Prescribers are required to complete a patient status form every 6 months for about 4 years

Table 24.3 Goals of the Alemtuzumab risk evaluation and mitigation strategy (REMS).

The goal of the LEMTRADA REMS is to mitigate the risks of autoimmune conditions, infusion reactions, stroke, and malignancies associated with LEMTRADA by:

1) Helping to ensure informed decisions about the safe use of LEMTRADA by:
   - Informing patients about the serious risks of autoimmune conditions, infusion reactions, and malignancies with LEMTRADA and the need for baseline and periodic monitoring; and
   - Informing healthcare providers about the serious risks of autoimmune conditions, infusion reactions, stroke, and malignancies with LEMTRADA, the need to counsel patients, and the need for baseline and periodic monitoring.
2) Helping to ensure the safe use of LEMTRADA by:
   - Ensuring that only certified prescribers prescribe LEMTRADA;
   - Ensuring that LEMTRADA is dispensed only in certain healthcare settings, by certified pharmacies, and certified infusion sites, which have on-site access to equipment and personnel trained to manage infusion reactions; and
   - Ensuring that only enrolled and authorized patients receive LEMTRADA;
   - Ensuring that certified prescribers submit documentation of periodic monitoring of patients who receive LEMTRADA to identify autoimmune conditions and malignancies.

after the final dose for each patient to capture adverse events of interest. The purpose of this is to further characterize the long-term safety of alemtuzmab.

The final component of the alemtuzumab REMS is an implementation system, which describes how operational elements and responsibilities specified in the REMS will be implemented by the manufacturer. The manufacturer is also responsible for conducting an assessment of whether the alemtuzumab REMS is meeting its risk mitigation goals. Assessments are required every 6 months from the date of approval of the REMS for 1 year, and annually thereafter. The assessment plan includes primarily process indicators, including outreach efforts by the manufacturer; the numbers of certified prescribers and infusion centers; results of assessments of the knowledge of prescribers, other healthcare providers, and patients about the risks and required safe use conditions; deviations from program requirements, for example administering alemtuzumab to patients who have not been enrolled in the program; information regarding any unintended consequences such as unintended delays in treatment; and an evaluation of whether appropriate patient monitoring has occurred, including reports of adverse events of interest.

The alemtuzumab REMS includes features of a modern pharmaceutical risk management program. First, the goal of this risk mitigation strategy is not prevention of the serious adverse events associated with alemtuzumab, but rather the program was instituted to ensure that the benefits of the drug outweigh its risks. To accomplish this, the FDA required that a rigorous safety program be implemented to mitigate the severity of the adverse events (i.e., improving outcomes in patients who may experience an infusion-related reaction) and early detection of serious autoimmune disorders so that patients can be appropriately managed.

Risk Management Example in the EU: Bupropion/Naltrexone (Mysimba)

Bupropion/naltrexone (with the brand name Mysimba and referred to as Mysimba throughout) was authorized in the EU in March 2015. It is a medicine used together with diet and exercise to treat obesity, which is defined as a BMI (body-mass index, a measure of weight relative to height) of 30 or above; it can also be given to very overweight patients who have weight-related complications.

The main safety and tolerability concerns identified with Mysimba were related to central nervous system and gastrointestinal adverse events, and uncertainties with regard to cardiovascular outcomes in the longer term. The manufacturer submitted an EU-RMP which included a risk minimization plan.

The CHMP decided that there was a need for both additional pharmacovigilance activities and risk minimization activities.

Additional pharmacovigilance activities for Mysimba

Apart from routine pharmacovigilance (i.e., a spontaneous reporting system), the manufacturer agreed to request the results of two Phase IV randomized clinical trials to evaluate the effect of Mysimba on the occurrence of major adverse cardiovascular events in overweight and obese subjects with high cardiovascular risk. Also, a drug utilization study and a physician survey were requested to evaluate how Mysimba is used in real-world medical practice because of the potential for off-label use.

Risk minimization activities: Mysimba

The most important safety concerns for Mysimba are seizures, suicidality in patients with depression, and off-label use. The main risk minimization strategy was to prevent patients with an increased risk of these adverse drug reactions from being prescribed Mysimba.

This translates into risk minimization activities to ensure the following:

- Mysimba is used within the SmPC indication.
- Patients with contraindications are not prescribed Mysimba.
- Physicians understand which patients (without contraindications) are likely to have additional risk factors for safety concerns and may need additional monitoring/counseling.
- The benefit/risk balance is positive at the individual patient level.

To reinforce the indication and warning and precautions in the SmPC and prevent off-label use, a physician prescribing checklist was proposed to guide physicians in prescribing appropriately. The checklist reminded the physician not to prescribe Mysimba in the presence of:

- Uncontrolled hypertension
- Current seizure disorder or a history of seizures or known central nervous system tumor
- Current or previous diagnosis of bulimia or anorexia nervosa
- Currently dependent on chronic opioids or opiate agonists (e.g., methadone), or patients in acute opiate withdrawal
- Undergoing acute alcohol or benzodiazepine withdrawal
- Concomitant treatment containing bupropion or naltrexone
- History of bipolar disorder
- Receiving concomitant administration of monoamine oxidase inhibitors
- End-stage renal failure or severe renal impairment

Also the checklist contained information to carefully evaluate the benefits and risk of treatment in the presence of the following conditions:

- Mild or moderate renal impairment
- Controlled hypertension
- Angina or recent history of myocardial infarction
- History of mania
- Suicidal ideation
- Depression
- Risk factors for seizures

In postmarketing, based on a review of the clinical trials and postmarketing reports of serious and nonserious cases of hepatotoxicity, there was evidence to suggest a possible causal relationship between hepatotoxicity and the use of Mysimba. Nonserious cases were reported to result in an elevation of liver enzymes. In addition, drug-induced liver injury (DILI) was described in clinical trials.

Therefore, in view of the data presented, it was considered that changes to the product information of all medicinal products containing the combination product naltrexone/bupropion were warranted, although there was no need to change the additional risk minimization in place until the results of the drug utilization and physician survey were provided.

The Future

Managing the risk of medicinal products is an evolving area involving multiple stakeholders in the complex medication use system. In this section we describe possible areas for future refinements of the current risk mitigation systems that are in place.

One critical area for future development is to continue to improve the way risk mitigation activities are being implemented. Many of the risk mitigation tools have relied in whole or in part on communicating a risk associated with a medicinal product to increase stakeholders' awareness and knowledge. The goal is that awareness of a particular risk will impart knowledge and influence prescribing practices of the medicinal product or how practitioners will monitor patients once treatment with the product has begun. The communication can be in the form of a one-time letter or ongoing communication to practitioners, a formal training program

that practitioners must undertake in order to prescribe the medicinal product, or an agreement by the practitioner that they will undertake certain conditions of safe use (e.g., to perform laboratory testing on a prespecified periodic basis). Communication is necessary, but delivery may be more effective if done in real time rather than weeks or months prior to prescribing the medicinal product.

Risk management plans are designed to work within a complex medication use system. The complex systems were developed before the advent of contemporary risk management planning efforts. A current challenge for risk management systems is that they be developed in ways that can integrate with minimal difficulty into the current medication use systems and in a manner that is least burdensome to healthcare providers. Risk mitigation strategies that require documentation of safe use conditions need refinement. A future challenge is to develop a quality systems approach to implementing this type of risk mitigation strategy in a manner that is seamless for practitioners and within the scope of their usual clinical practice. The design of this type of mitigation strategy can be challenging, because the medication use process is complex. Furthermore, the medication use process differs across various healthcare settings, which makes it additionally challenging to implement mitigation strategies with the same goals across the various settings. Evaluation of its impact will also be challenging and difficult to answer, and will require the integration of many disciplines, including pharmacoepidemiology.

Another area for future refinement is to continue to gather evidence of the impact of various risk mitigation strategies, particularly on safety or health outcomes. Measurement of the impact is important, because it allows policymakers and other stakeholders to determine if the goals of the strategy are being met. Much of the evidence collected to date has been on process indicators such as whether particular safety messages have reached the intended stakeholder, whether it imparted the knowledge necessary to use the medicinal product safely, and whether that knowledge resulted in a change in the clinical practice of the practitioner (i.e., whether it resulted in the desired behavior by the stakeholder). Disciplines including social science and health service research have made significant progress in advancing our understanding of those types of outcomes. As noted earlier, the most important outcome measures to assess are the specific health outcomes of interest. If process outcomes and behavioral outcomes are met, but there is no impact on health outcomes or the impact on health outcomes is unknown, risk management strategies may need to be reconsidered. The optimal way to measure the impact of risk management strategies on safety or health outcomes remains a challenge, and is an area in which pharmacoepidemiology plays a crucial role.

Pharmacoepidemiology is integral to the measurement of specific health outcomes. The challenges for this field include developing models to relate risk management strategies to health outcomes, as well as ways to identify the contribution of individual components of the strategy to the overall outcome. Evaluations of the risk mitigation strategy should include an assessment of any negative consequences to implementing a risk management strategy. For example, if restricted distribution is put in place for a certain medicine that requires documentation that a laboratory test was conducted prior to dispensing the medicine to the patient, does this strategy, because of the real or perceived burdens associated with its implementation, result in patients not receiving the medicine even if it is the most appropriate medicine for that patient? This assessment should be followed by an evaluation of the impact of the risk mitigation strategy on stakeholders, including prescribers, pharmacists, and patients, to ensure the strategy improves the safe use of the product, but does so in the most efficient manner possible.

As risk management planning is evolving in multiple countries and regions, there is considerable interest in international harmonization of these efforts. At this time, there are many challenges that must be overcome if harmonization is to become a reality. First, the diversity of healthcare systems and medication use systems from one country to the next limits the degree to which identical, or even similar, individual risk mitigation plans can be put in place across several countries. Second, because of the differences in risk mitigation systems that can be put in place across countries, it would be challenging to develop a common risk management plan that manufacturers could submit to all regulatory authorities. The differences in risk management activities across countries and regions, however, create a natural opportunity for stakeholders to determine the relative impact of different approaches to risk management.

References

1 US Food and Drug Administration. *Guidance for Industry: Development and Use of Risk Minimization Action Plans*. Silverspring, MD: FDA, 2005.
2 Directive 2001/83 EC, Article 1 (28b).
3 National Coordinating Council for Medication Error Reporting and Prevention. About medication errors. http://www.nccmerp.org/about-medication-errors (accessed November 2017).
4 European Medicines Agency. Good pharmacovigilance practices. https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/good-pharmacovigilance-practices (accessed July 2019).
5 Massachusetts Technology Collaborative and NEHI. *Saving Lives, Saving Money: The Imperative for Computerized Physician Order Entry in Massachusetts Hospitals*. Westborough, MA: Massachusetts Technology Collaborative, 2008.
6 Center for Information Technology Leadership. The value of computerized provider order entry in ambulatory settings. https://healthit.ahrq.gov/sites/default/files/docs/page/CITL%20ACPOE%20Executive%20Preview_0.pdf (accessed November 2017).
7 Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human: Building a Safer Health System*. Washington, DC: Institute of Medicine, National Academies Press, 2000.
8 NEHI. Preventing medication errors: a $21 billion opportunity. http://www.nehi.net/bendthecurve/sup/documents/Medication_Errors_%20Brief.pdf (accessed November 2017).
9 Newbould V, Le Meur S, Goedecke T, Kurz X. Medication errors: a characterisation of spontaneously reported cases in EudraVigilance. *Drug Safety* 2017; **40**: 1241–48.
10 Blossom DB, Kallen AJ, Patel PR, *et al*. Outbreak of adverse reactions associated with contaminated heparin. *N Engl J Med* 2008; **359**: 2674–84.
11 McMahon AW, Pratt RG, Hammad TA, *et al*. Description of hypersensitivity adverse events following administration of heparin that was potentially contaminated with oversulfated chrondroitin sulfate in early 2008. *Pharmacoepidemiol Drug Saf* 2010; **19**: 921–33.
12 Pinnow E, Amr S, Bentzen SM, *et al*. Postmarket safety outcomes for new molecular entity (NME) drugs approved by the Food and Drug Administration between 2002 and 2014. *Clin Pharmacol Ther* 2018; **104**(2): 390–400.
13 Larson AM, Polson J, Fontan RJ, *et al*.; Acute Liver Failure Study Group. Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology* 2005; **42**: 1364–72.

14 Goldkind L, Laine L. A systematic review of NSAIDs withdrawn from the market due to hepatotoxicity: lessons learned from the bromfenac experience. *Pharmacoepidemiol Drug Saf* 2006; **15**: 213–20.

15 US Food and Drug Administration. *Guidance for Industry. Safety Considerations for Product Design to Minimize Medication Errors*. Silversprings, MD: FDA, 2016.

16 European Medicines Agency. Good practice guide on risk minimisation and prevention of medication errors; Annex 1 – Potential sources of medication error in medicinal product design. 2015. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2015/11/WC500196981.pdf (accessed November 2017)

17 21 CFR 208.

18 Ovation Pharmaceuticals, Inc. Advisory Committee briefing document. Sabril® (vigabatrin) tablet and powder for oral solution for adjunctive treatment of refractory complex partial seizures in adults (NDA 20-427), for monotherapy treatment of infantile spasms (NDA 22-006). Peripheral and Central Nervous System Advisory Committee 7–8 January 2009. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2009/022006s000_MedR.pdf (accessed May 2019).

19 FDA Sentinel System. https://www.sentinelsystem.org/sentinel/about (accessed November 2017).

20 Cocoros N, Haynes H, Tu C, *et al*. Identification of name confusion medication errors in the Sentinel System. Presented at the International Conference on Pharmacoepidemiology and Therapeutic Risk Management, Montreal, Canada, August 2017.

21 National Asthma Education and Prevention Program's (NAEPP). Expert Panel Report 3 (EPR-3). https://www.nhlbi.nih.gov/files/docs/guidelines/asthsumm.pdf (accessed July 2011).

22 Friedman H, Wilcox T, Reardon G, Crespi S, Yawn BP. A retrospective study of the use fluticasone proprionate/salmeterol combination as initial asthma controller therapy in a commercially insured population. *Clin Ther* 2008; **30**: 1908–17.

23 Chowdhury BA, Dal Pan G. The FDA and safe use of long-acting beta-agonists in the treatment of asthma. *N Engl J Med* 2010; **362**: 1169–71.

24 Nourjah P, Ahmad SR, Karwoski C, Willy M. Estimates of acetaminophen (Paracetamol)-associated overdoses in the United States. *Pharmacoepidemiol Drug Saf* 2006; **15**: 398–405.

25 Statutory warnings for all medicines containing paracetamol. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/328330/Statutory_warnings_for_all_medicines_containing_paracetamol.pdf (accessed November 2017)

26 Morgan O, Griffiths C, Majeed A. Impact of paracetamol pack size restrictions on poisoning from paracetamol in England and Wales: an observational study. *Journal of Public Health* 2005; **27**: 19–24.

27 FDA Drug Safety Communication: Prescription acetaminophen products to be limited to 325 mg per dosage unit; boxed warning will highlight potential for severe liver failure. https://www.fda.gov/Drugs/DrugSafety/ucm239821.htm (accessed November 2017).

28 Brinker A, Kornegay C, Nourjah P. Trends in adherence to a revised risk management program designed to decrease or eliminate isotretinoin-exposed pregnancies: evaluation of the Accutane SMART program. *Arch Dermatol* 2005; **141**: 563–9.

29 Joint Meeting of the Drug Safety and Risk Management Advisory Committee and the Dermatologic and Ophthalmic Drugs Advisory Committee, December 1, 2011. Background materials. https://wayback.archive-it.org/7993/20170403223855/https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/DrugSaf

etyandRiskManagementAdvisoryCommittee/ucm250295.htm (accessed November 2017).
30 Shrank W, Avorn J, Rolon C, Shekelle P. Effect of content and format of prescription drug labels on readability, understanding, and medication use: a systematic review. *Ann Pharamacother* 2007; **41**: 783–801.
31 Shrank W, Gleason P, Canning C, *et al*. Can improved prescription medication labeling influence adherence to chronic medications? An evaluation of the Target pharmacy label. *J Gen Intern Med* 2009; **24**: 570–8.
32 Public Law 110-85 (September 27, 2007). https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/html/PLAW-110publ85.htm (accessed November 2017).
33 US Food and Drug Administration. *Guidance for Industry: FDA's Application of Statutory Factors in Determining When a REMS Is Necessary*. Silversprings, MD: FDA, 2019.
34 Public Law 115-271 (October 24, 2018). https://www.congress.gov/bill/115th-congress/house-bill/6 (accessed July 22, 2019).
35 European Medicines Agency. Risk minimization strategy for high strength and fixed combination insulin products (EMA/686009/2014. 2015. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/risk-minimisation-strategy-high-strength-fixed-combination-insulin-products-addendum-good-practice_en.pdf (accessed July 2019).
36 US Prescribing Information for LEMTRADA® (alemtuzumab) injection, for intravenous use, 10/2017. https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/103948s5163s5164slbl.pdf (accessed July 2019).
37 Ruck T, Bittner S, Wiendl H, Meuth SG. Review Alemtuzumab in multiple sclerosis: mechanism of action and beyond. *Int J Mol Sci* 2015; **16**: 16414–39.
38 US Approved Lemtrada REMS. https://www.accessdata.fda.gov/scripts/cder/rems/index.cfm?event=IndvRemsDetails.page&REMS=340 (accessed July 2019).

25

Distributed Networks of Databases Analyzed Using Common Protocols and/or Common Data Models

Sengwee Toh[1], Nicole Pratt[2], Olaf Klungel[3], Joshua J. Gagne[4], and Robert W. Platt[5]

[1] *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA*
[2] *Quality Use of Medicines and Pharmacy Research Centre, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia*
[3] *Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, The Netherlands*
[4] *Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA*
[5] *Departments of Epidemiology, Biostatistics, and Occupational Health, and of Pediatrics, McGill University, Montreal, Quebec, Canada*

It is now common to analyze large and complex electronic healthcare databases – created as part of regular business operations or routine clinical care – to assess the safety and effectiveness of medical products. Pharmacoepidemiologic studies have traditionally analyzed information available in single databases. However, single-database studies may not be sufficient to answer certain clinical questions, especially when the exposure or outcome is rare, when the goal is to study the treatment effect in specific subgroups, or when the objective is to identify a sufficiently large number of exposed patients within a relatively short time window (e.g., during the early months following the approval of a new medical product).

Thanks to the increase in the number of data sources, and the improvement in the quality of and ease of access to these data sources, multidatabase studies are now feasible and ubiquitous in pharmacoepidemiologic research. There are several ways to conduct multidatabase studies. An intuitive approach is to pool the databases or the derived analytic datasets centrally for analysis. However, centralized pooling of databases that contain detailed individual-level data is not always possible for several reasons, including concerns about patient privacy, data security, unauthorized uses of data, and potential disclosures of sensitive institutional or business information. A distributed approach, in which databases are not combined centrally but rather stored in different physical locations under the direct control of the participating sites, is becoming increasingly preferred.

In this chapter, we describe the design, development, implementation, strengths, and challenges of distributed data networks (DDNs). We begin with a brief description of select DDNs in pharmacoepidemiology. We then discuss the types of research questions that DDNs are designed to address. We examine the methodologic and data issues unique to DDNs, and progress that has been accomplished to address these issues. We conclude with a discussion about some of the future directions for DDNs.

Here we first define three key terms that are central to the chapter:

- *Distributed data network*: two or more data sources stored in different physical locations

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

under the direct control of the participating data partners.
- *Common data model (CDM)*: a data model that generally includes a set of standardized data files and variables, adopted by all data sources participating in a DDN.
- *Common protocol*: a study protocol that typically includes detailed description of key design and analytic parameters of the study, implemented by all data sources participating in a DDN study.

Examples of Distributed Data Networks in Pharmacoepidemiology

DDNs have been in existence for more than 20 years. Drawing on more than two decades of experience, the DDNs today are more sustainable, efficient, and diverse. This section briefly describes a number of DDNs designed to conduct pharmacoepidemiologic research, medical product safety surveillance, or comparative effectiveness research using electronic health data collected as part of routine healthcare delivery. Table 25.1 provides a summary of their key characteristics.

Asian Pharmacoepidemiology Network (AsPEN)

Established in 2008, AsPEN is a multinational research network formed to provide a mechanism to support the conduct of pharmacoepidemiologic research and to facilitate more rapid identification and validation of emerging safety issues among the Asia-Pacific countries [1,8]. AsPEN is a collaboration of 8 countries involving 12 databases formalized as a Special Interest Group of the International Society for Pharmacoepidemiology. It has piloted a number of approaches to conduct distributed studies including a common protocol, a standard analytic program [9–12], and translation to a CDM developed by the Observational Medical Outcomes Partnership (OMOP; see later) [13].

Canadian Network for Observational Drug Effect Studies (CNODES)

CNODES is a distributed network of Canadian research teams designed to provide evidence to Canadian stakeholders, in particular Health Canada, on drug safety in the Canadian context [2]. It is one of four collaborating centers supported by the Drug Safety and Effectiveness Network (DSEN) of the Canadian Institutes of Health Research. Queries from stakeholders are prioritized by DSEN head office and sent to the CNODES coordinating center. CNODES teams conduct analyses of a distributed network of Canadian and international databases, and report both to stakeholders and via published literature.

Health Care Systems Research Network (HCSRN)

Established in 1994, HCSRN (formerly known as the Health Maintenance Organization Research Network, HMORN) is a consortium of 18 integrated delivery systems and health plans designed to facilitate multidatabase collaborative research [3]. Compared to other DDNs, HCSRN is unique because it is not created to address specific clinical or research questions. Instead, it is "multipurpose" and supports a wide range of research and surveillance activities. HCSRN is the foundation on which several large-scale collaborative projects are built and maintained, including the Vaccine Safety Datalink (VSD; more later) [14], the Cancer Research Network [15], the Mental Health Research Network [16], and the Cardiovascular Research Network [17]. HCSRN is often considered one of the best examples of a sustained DDN [3]. Its distributed network architecture and CDM (known as the Virtual Data

Table 25.1 Select examples of distributed data networks in pharmacoepidemiology.

	AsPEN [1]	CNODES [2]	HCSRN [3]	PCORnet [4]	PROTECT [5]	Sentinel [6]	VSD [7]
Number of data partners	12	9	18	>80	14	18	9
Total population	220 million	35 million (Canada)	16 million	100 million	100 million	293 million	9 million
Type of data	Claims	Claims, EHRs	Claims, EHRs	Claims, EHRs	Claims, EHRs	Claims, EHRs	EHRs
Geography	Asia-Pacific	Canada, US, and UK	US and Israel	US	European Union	US	US
Funding source	None	CIHR	Various	PCORI	IMI	FDA	CDC
Primary mission	Medication safety research	Medical product safety research	Multipurpose	Patient-centered outcomes research	Medication safety research	Medical product safety surveillance	Vaccine safety surveillance
Common data model	Yes	Yes	Yes	Yes	No	Yes	Yes
Common study protocol	No	Yes	Yes	Yes	Yes	Yes	Yes
Common statistical analysis plan	Yes	Yes	Yes	Yes	Yes	Yes	Yes

AsPEN, Asian Pharmacoepidemiology Network; CDC, Centers for Disease Control and Prevention; CIHR, Canadian Institutes of Health Research; CNODES, Canadian Network for Observational Drug Effect Studies; EHRs, electronic health records; FDA, Food and Drug Administration; HCSRN, Health Care Systems Research Network; IMI, Innovative Medicines Initiative; PCORI, Patient-Centered Outcomes Research Institute; PCORnet, National Patient-Centered Clinical Research Network; PROTECT, Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium; VSD, Vaccine Safety Datalink.

Warehouse) [18] have been adopted by other DDNs, such as the Sentinel system (see later).

National Patient-Centered Clinical Research Network (PCORnet)

Launched in 2013, PCORnet is a network of networks that includes 13 Clinical Data Research Networks, 20 People-Powered Research Networks, 2 Health Plan Research Networks, and a coordinating center in the US [4]. The Clinical Data Research Networks are primarily comprised of healthcare delivery systems, the Health Plan Research Networks are shepherded by two national insurers, and the People-Powered Research Networks are mainly led by patient and caregiver organizations. PCORnet is designed to support both randomized trials and observational studies. It currently includes electronic health record (EHR) or administrative claims data from more than 100 million individuals and has access to over 40 million patients who could be recruited into pragmatic clinical trials.

Pharmacoepidemiologic Research on Outcomes of Therapeutics by a European Consortium (PROTECT)

Initiated in 2009 and ended in 2015, PROTECT was a joint undertaking by the European Union (EU) and pharmaceutical industry as part of the Innovative Medicines Initiative [5]. Its 35 partners, including academics, regulators, small and medium enterprises, and member companies of the European Federation of Pharmaceuticals Industries and Associations, were coordinated by the European Medicines Agency. The overall objective of PROTECT was to address limitations of current methods in the field of pharmacoepidemiology and pharmacovigilance. As part of this work, a network of electronic healthcare databases was established to conduct multicountry, multidatabase, drug safety studies. Currently, several former public partners of PROTECT are continuing their collaboration with additional public partners in the European Research Network for Pharmacoepidemiology and Pharmacovigilance, allowing access to a broad variety of datasets (general practice, hospital pharmacy/laboratory, pharmacy, hospitalization, claims, questionnaires, and biological samples) covering six EU countries (Spain, UK, Italy, the Netherlands, Denmark, and France) and records from approximately 100 million active patients to address various research questions.

Sentinel System

The Sentinel system is funded by the US Food and Drug Administration (FDA) as a national medical product surveillance system mandated by the US Congress in the FDA Amendments Act of 2007 [6,19,20]. Initiated as a pilot program called Mini-Sentinel in 2009 [21], the system includes a distributed network of 18 data partners that provide access to administrative claims and EHR information from over 290 million cumulative patient identifiers. A feature of Sentinel is its ability to conduct rapid descriptive and inferential analysis using preprogrammed, pretested, and customizable analytic tools [22–25]. These analytic tools and other Sentinel-related materials (e.g., protocols, reports, data model) are all available in the public domain.

Vaccine Safety Datalink (VSD)

Funded by the US Centers for Disease Control and Prevention since 1990, VSD monitors the safety of vaccines using EHR databases from a network of nine delivery systems and health plans [7,14,26,27]. While VSD initially used a centralized data model in which the data partners submitted de-identified analytic datasets for centralized analysis, it switched to a more sustainable DDN model in 2001 [28]. A unique feature of VSD is its ability to provide near

real-time surveillance of vaccine safety. Researchers apply study design and statistical methods appropriate for sequential surveillance to analyze weekly updated data. The analysis takes into account data lag and incompleteness due to frequent refreshes of the data [29,30].

Others

OMOP was a partnership between the FDA and the Pharmaceutical Research and Manufacturers of America established to inform the appropriate use of electronic healthcare databases for studying the effects of medical products [13]. In achieving that goal, OMOP initiated a series of experiments to investigate the feasibility and validity of conducting fully automated assessments of medical product safety [31,32]. The initiative also created a CDM and a suite of CDM-compatible analytic tools. The OMOP experiment was completed in 2013, but much of its collaborative work continues within Observational Health Data Sciences and Informatics (OHDSI) [33].

The "Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge" (EU-ADR) project was launched in 2008 with the aim to leverage information from various EHR databases in Europe to produce a computerized integrated system for the early detection of drug safety signals [34]. The same approach and some of the same databases were used in study-specific networks, including the Safety of Non-steroidal Anti-inflammatory Drugs (SOS) [35], Arrhythmogenic Potential of Drugs (ARITMO) [26], Safety Evaluation of Adverse Reactions in Diabetes (SAFEGUARD) [37], Global Research in Paediatrics (GRIP) [38], and Accelerated Development of Vaccine Benefit-risk Collaboration in Europe (ADVANCE) [39] projects. Apart from EU-ADR, these collaboratives have included several of the same EU datasets and developed study-specific CDMs.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

As with multidatabase studies that analyze centrally pooled information, DDN studies support analyses that cannot typically be done with one data source. Examples include assessments of rare exposures, rare outcomes, treatment effect heterogeneity in specific subpopulations, and surveillance of newly approved medical products (Table 25.2).

Assessment of Rare Exposures

Examples of rare exposures include, but are not limited to, drugs used to treat orphan diseases, defined as conditions that affect fewer than 200 000 individuals (US definition) or 5 in 10 000 individuals (EU definition). DDNs allow studies of drugs indicated for orphan diseases [59,60]. For example, several People-Powered Research Networks within PCORnet, such as the Phelan-McDermid Syndrome Data Network [61], are leveraging patient-generated information and the electronic health data within Clinical Data Research Networks to generate evidence about disease progression and treatments for these conditions.

Assessment of Rare Outcomes

An example of rare outcomes is Guillain-Barré syndrome, which occurs in 1–2 per 100 000 person-years [62]. An adequately powered study to examine the association between a vaccine and Guillain-Barré syndrome requires information from millions of individuals from multiple databases. For example, to assess the risk of Guillain-Barré syndrome following receipt of a quadrivalent human papillomavirus vaccine, a VSD study used six databases to identify males and females aged 9–26 years who received the vaccine from 2006 to 2015 [58]. One confirmed case of Guillain-Barré syndrome within 42 days

Table 25.2 Select examples of studies conducted within distributed data networks.

Network	Select studies
AsPEN	• Antipsychotic use and risk of acute hyperglycemia [9] • Thiazolidinedione use and risk of heart failure across ethnic groups [11] • Cardiac safety of methylphenidate among pediatric patients with ADHD [40]
CNODES	• Statin use and risk of acute kidney injury [41] • Incretin-based drug use and risk of heart failure [42] • Occurrence of pregnancy during isotretinoin therapy [43]
HCSRN	• Lipid-lowering drug use and risk of rhabdomyolysis [44] • Prenatal antidepressant exposure and risks of congenital malformations [45] • ADHD medication exposure and risk of serious cardiovascular events [46]
PCORnet	• Aspirin dosing and secondary prevention of atherosclerotic cardiovascular disease [47] • Antibiotic use and weight outcomes in children [48] • Long-term benefits and risks of bariatric procedures [49]
PROTECT	• Antibiotic use and risk of acute liver injury [50] • Antidepressant use and risk of hip fracture [51] • Antiepileptic drug use and risk of suicidality [52]
Sentinel	• Antihyperglycemic use and risk of acute myocardial infarction [53] • Dabigatran use and risks of bleeding and cardiovascular events [54] • Rotavirus vaccination and risk of intussusception [55]
VSD	• Thimerosal exposure and risks of neuropsychological outcomes [56] • Safety of H1N1 and seasonal influenza vaccines [57] • Quadrivalent human papillomavirus vaccination and risk of Guillain-Barré syndrome [58]

ADHD, Attention-deficit hyperactivity disorder; AsPEN, Asian Pharmacoepidemiology Network; CNODES, Canadian Network for Observational Drug Effect Studies; HCSRN, Health Care Systems Research Network; PCORnet, National Patient-Centered Clinical Research Network; PROTECT, Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium; VSD, Vaccine Safety Datalink.

following vaccination was confirmed among over 2.7 million vaccinees.

Assessment of Treatment Effect Heterogeneity

Certain treatments may have different effectiveness or safety profiles in patients with specific characteristics. Combining information from multiple databases allows researchers to have enough sample sizes in subsets of populations (e.g., children, the elderly, individuals with a history of heart failure). Results from studies of databases that cover demographically and geographically diverse populations also provide better generalizability. For example, AsPEN conducted a study to determine whether the risk of edema or heart failure associated with thiazolidinediones was different between Caucasian and Asian populations, potentially due to differences in the prevalence of metabolizing enzymes in these ethnic groups [11]. In another example, a Sentinel study examined the associations between antihyperglycemic treatments and risk of hospitalized heart failure among diabetes patients with and without a history of cardiovascular disease [63]. The European Research Network for Pharmacoepidemiology and Pharmacovigilance

is currently investigating the risk of major bleeding associated with direct oral anticoagulants in targeted clinical and demographic subgroups for which variations in plasma concentrations might affect the safety of the products [64].

Postmarket Surveillance of Newly Approved Medical Products

An increasingly common scenario that warrants the use of multiple databases is postmarket surveillance of the safety of newly approved medical products. The goal is to monitor new medical products as postmarket experiences of their use accrue in routine clinical practice. The number of users in a single database is usually low in the early postapproval phase, so multiple databases are required to support an informative analysis. Prospective, sequential analysis of cumulating data can be addressed by appropriate statistical techniques [65–67]. For example, the Sentinel system has leveraged its DDN to complete prospective postmarket surveillance of two newly approved medical products and select health outcomes. One assessed the association between saxagliptin (an oral antihyperglycemic agent) and acute myocardial infarction following the approval of the drug in 2009 [53]. The other examined the associations between rivaroxaban and ischemic stroke, intracranial hemorrhage, and major gastrointestinal bleeding after the oral anticoagulation drug was approved in 2011 [68].

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

Pooling of individual-level data, often in a format that is stripped of direct identifiers, had traditionally been the default approach used in multidatabase analyses. Sharing of de-identified individual-level data is generally feasible in the presence of proper governance, appropriate data use or sharing agreements, and established collaborative relationships [69]. However, concerns about patient privacy and confidentiality, unauthorized uses of transferred data, and unintended disclosures of sensitive corporate or institutional information have made data sharing increasingly more challenging in practice, specifically in newly formed collaborations or projects that perform a large number of studies [70–72]. Contractual agreements between health plans and some of their members may further restrict sharing of individual-level information with other entities for secondary purposes such as research.

A DDN architecture addresses some of the concerns associated with pooling of individual-level data [70–72]. A typical DDN generally has the following features:

- There is one or more coordinating center(s).
- Data partners maintain physical control of their data.
- Data partners have the ability to review and approve each data request.
- Data partners have the ability to review the output before sharing it with the requester.
- Data partners can opt out of any data request at any time.

These features offer data partners more autonomy in multidatabase studies. They allow data partners to evaluate their ability or willingness to share their data with the requester at various steps of the request. More importantly, a DDN approach keeps the data close to the individuals who know the data best. The data partners can advise on the appropriate use of the data and help investigate data anomalies and interpret findings.

Challenges in Distributed Data Networks

DDNs also come with challenges, some of which are common across all multidatabase studies and others unique to the distributed environment.

There could be heterogeneity in data quality, data completeness, coding system, and patient population. Neither a centralized nor a distributed system is immune to these issues, but they can be more difficult to identify or diagnose in a distributed environment. Compared to single-database studies, multidatabase studies often involve additional administrative and governance issues, such as the need for multicenter ethics review and data-sharing agreements. These issues may sometimes (but not always) be more resource intensive in DDNs than a centralized system, depending on the type of analysis and information shared. There may be a need for more frequent communications between the coordinating center and participating sites in DDNs, which can create delays. In multinational DDNs, the participating data sources may have different languages, different availability of medical products, wider variation in clinical practice and drug utilization, and different responses to medical products due to ethnic or genetic variations. The conduct of the statistical analysis is generally more complicated in DDNs, because individuals responsible for the analysis do not have full access to the source data from all the participating sites. Although DDNs increase overall sample size, the larger sample size does not necessarily help improve control for confounding, since confounding control typically occurs separately within each site.

Currently Available Solutions

To facilitate the conduct of studies, existing DDNs organize themselves differently based on their resources, needs, expertise, and data infrastructure. On one end of the spectrum are DDNs that employ a common protocol and a CDM approach. At the other end of the spectrum are DDNs that have neither a common protocol nor a CDM. There are also DDNs that adopt a common protocol approach without a CDM. These options lie on a continuum and do not represent all the possible scenarios. For example, a DDN can develop a CDM for some of its data partners but not the others. Each of these options has its unique strengths and limitations (Table 25.3). However, some offer clear advantages over the others in many scenarios. In particular, a DDN that has neither a CDM nor a common protocol approach is typically less efficient than the other systems. Table 25.1 summarizes the approaches employed by some of the DDNs.

Common Data Model, with or without a Common Protocol

Some DDNs have all participating data partners convert their source data into standardized data formats, often known as a CDM. The CDM specifies a uniform data file structure and data element naming conventions and definitions across all databases. There are several CDMs in use, including for Sentinel [72], PCORnet [73], HCSRN [18], and OMOP [74]. The first versions of the Sentinel and OMOP CDMs were modeled in part on the HCSRN CDM, and the Sentinel CDM served as the backbone of the PCORnet CDM. CNODES has implemented the Sentinel CDM in four databases, is working to implement it network-wide, and has initiated queries using the CDM.

There is a general misconception that a CDM is a "lowest common denominator" approach, which reduces the data elements in a DDN to only variables common across all databases. In reality, data partners with more information can populate additional tables or variables for use in specific studies. For example, both the HCSRN and Sentinel CDMs allow data partners with clinical information from EHRs to populate additional tables on vital signs and laboratory test results [18,72].

However, certain information may be lost during the standardization process. This may occur when the data elements are available in multiple coding systems and researchers attempt to map across these systems. For example, US databases primarily use the International

Table 25.3 Strengths and limitations of various structures of distributed data networks.

Common data model	**Yes**	**Yes**	**No**	**No**
Common protocol	**Yes**	**No**	**Yes**	**No**
Upfront data infrastructure investment	Substantial	Substantial	Minimal	Minimal
Site-specific statistical programming effort	Minimal to moderate	Minimal to moderate	Moderate to substantial	Moderate to substantial
Ability to develop preparameterized, reusable tools	Yes	Yes	Limited	Limited
Ability to assess database heterogeneity	Yes	Yes	Yes	Yes
Ability to perform analysis tailored to individual databases	3 or 4 (worst)	2 or 3	2 or 3	1 (best)
Ability to ensure consistent analysis across databases	1 (best)	2 or 3	2 or 3	4 (worst)
Study-specific data management and cleaning	Minimal to moderate	Minimal to moderate	Substantial	Substantial
Speed of study-specific analysis	1 (fastest)	2 or 3	2 or 3	4 (slowest)
Marginal cost per study	1 (lowest)	2 or 3	2 or 3	4 (highest)
Reproducibility/validation across sites	1 (best)	3	2	4 (worst)

Classification of Diseases, 9th or 10th Revision, Clinical Modification (ICD-9-CM or ICD-10-CM) coding systems to record diagnoses, while the general practice databases in the UK document diagnoses using Read codes. It is possible to map and standardize these coding systems, but doing so may lead to some information loss or misclassification. In the presence of multiple coding systems, it is still possible to develop a CDM while preserving the fidelity or granularity of the source information for use in actual studies. In the diagnosis example earlier, the CDM will have a variable that contains the specific diagnosis codes (G30z.00 or 410.00) and an additional variable that indicates the code type (Read or ICD-9-CM). DDN studies that analyze against the CDM can then use the two variables together to define the study parameters. In the presence of multiple coding systems, input from researchers and others familiar with the data is required, either during the mapping process or when conducting a study.

DDNs with a CDM almost always conduct their studies with a common protocol (more later). Analyzing CDM-backed databases with a common protocol allows study-specific data checking, management, and analysis to be done via identical computer programs that can be developed and beta-tested by a smaller group of individuals. This helps reduce programming burden at sites, minimizes opportunities for errors across participating sites, and ensures consistent analysis across databases. However, the centrally developed computer programs have to accommodate differences in computing environments (e.g., different operating systems, software versions) to allow successful execution across all participating sites. In addition, a coding error that occurs in a centrally developed program will have an impact on all sites.

On the rare occasions that a common protocol is not developed in the presence of a CDM, the data partners presumably would have more flexibility in answering the study question. This would allow certain data partners that have more data elements in the CDM to include them in their analysis. However, it is worth noting that a common protocol can be developed in a way that also allows database-specific analysis, for example through a semi data-adaptive approach like high-dimensional propensity score analysis, in which the propensity score is built individually in each database using available information rather than using a common set of variables [23].

Common Protocol, with or without a Common Data Model

In a common protocol approach, a protocol is developed, often collaboratively among participating sites, for implementation across the DDN. As already discussed, employing a common protocol approach in the presence of a CDM generally allows the study to be conducted more efficiently, as programming burden is limited to one site rather than having each site develop *de novo* code. In the absence of a CDM, the common protocol is generally less prescriptive, to allow data partners to define the measurement of exposure, outcome, covariates, and other study parameters based on the information available in their databases. For example, in a study of rivaroxaban, the protocol will specify the exposure of interest, but each data source will identify rivaroxaban exposure based on its coding system (e.g., National Drug Codes or Anatomical Therapeutic Chemical Classification System).

As individuals who are most familiar with the data are actively involved in the implementation of the protocol, the study can accommodate the differences in coding practices, data quality and completeness, and other idiosyncratic issues associated with each database. The disadvantage of this approach is that it can lead to variations in the interpretation of the protocol, which may artificially inflate the heterogeneity across sites or affect the robustness of results. In addition, each site is required to have adequate

programming resources to conduct its own analysis. Coordination across the DDN during the study can be intensive in order to resolve any discrepancies and to ensure consistent interpretation of the analysis plan. However, this can be accomplished by using a detailed statistical analysis plan and a phased analysis, in which, for example, analyses are reviewed at various stages in the process (e.g., after baseline tables are populated, after propensity scores are estimated). Protocol refinements and modifications, as well as site-specific amendments (should initial analyses uncover differences in prescription patterns across sites), can help eliminate or explain discrepancies between sites. Some DDNs employ a blinding procedure to mask the results from participating data sources to facilitate more objective assessment of heterogeneity across sites. These processes to improve consistency require substantial time investment by analytic personnel. The common protocol approach can be particularly onerous for DDNs that do not share a common language or coding system across data partners.

The Necessity of Having a Common Data Model

Creating a CDM is a substantial undertaking. It requires considerable upfront investment on data infrastructure, in particular the extraction, transformation, and loading of the source data to a CDM. Additionally, the ongoing maintenance of the CDM can be burdensome, particularly as new versions of the CDM are needed, which can occur when there are new analytic requirements or changes in certain data elements. Each site must also routinely convert its new data into the CDM. It is generally easier to create a CDM for databases that contain the same type of information and coding system (e.g., claims data coded in the ICD-10-CM system). Developing a CDM for disparate data sources (e.g., claims databases and EHR databases) or databases with different coding systems is more challenging, but is possible through the use of mapping algorithms. However, as discussed earlier, there may be information loss or misclassification if mapping is required.

In general, it will be worthwhile to develop a CDM if the DDN is designed to conduct multiple studies. It may also be useful to convert the source data into a specific CDM to leverage available software or tools that are compatible with the CDM. For example, Sentinel, PCORnet, and OHDSI have developed a suite of analytic tools that can be executed within databases that use their CDMs. From the scientific validity perspective, the amount of data management, quality assurance, and harmonization for a given multidatabase study is similar regardless of the data network architecture. The CDM approach spends more resources upfront on data harmonization and quality assurance, so that downstream studies can be done more efficiently. However, the cost of establishing and maintaining the CDM is the same for 1 study as it is for 100 studies. The CDM achieves an economy of scale when the number of studies supported by the CDM is sufficiently large. In principle, the marginal cost of conducting a study is lower in DDNs with a CDM than in DDNs without a CDM when the number of studies is large.

A key consideration in developing a CDM is how much, if any, preprocessing of the information should be done upfront, and how much should be handled when conducting the study. Preserving the fidelity and granularity of the source information, as briefly discussed already, allows researchers of specific studies to determine the most appropriate study parameters. The extra time to execute the study, due to the additional deliberation, is generally worthwhile. Preprocessing the information upstream via mapping or creating specific constructs or concepts helps expedite the implementation of specific studies downstream, but may restrict researchers' ability to develop study parameters tailored to the studies. The approach taken by a given DDN depends on its missions, objectives,

and preferences. For example, the Sentinel system does minimal preprocessing of the data upfront, which allows the system to tailor its analysis to the FDA's regulatory questions. The OMOP CDM involves a significant amount of preprocessing, which allows researchers to streamline the conduct of their analysis using predefined variables and parameters. The pros and cons of these approaches, as well as their comparative performance, have been covered in the literature [75–78].

Methodological Advances

The defining feature of DDNs is that the data are stored locally under the direct control of participating data partners, and ideally only minimal necessary information is shared in each analysis. Traditionally, it had been necessary or preferred to share de-identified individual-level datasets for centralized analysis. With this approach, the participating sites send the analysis center an individual-level analytic dataset with distinct covariate information necessary for the analysis, yielding what is essentially a single centralized dataset after pooling. The confounders can be incorporated into the analysis through matching, stratification, restriction, regression, or weighting, and the data can be considered all together or stratified by contributing site [79,80]. Confounder summary scores (discussed shortly) can be estimated after centralizing the data. Although this approach offers the most analytic flexibility, it requires the most granular information among all the analytic options.

Recent methodological advances have expanded the data-sharing and analytic options, some of which require less granular information to perform the same type of analysis afforded by pooled individual-level data [81–85]. As a result, these newer methods may be preferred because they are more privacy protecting. Another feature of these new methods is that most or all the analyses will need to be specified *a priori*, or additional data requests may be required to obtain the additional information needed for *ad hoc* analyses. Although these are often seen as the limitations of these newer methods, they can also be considered strengths, because there is better transparency in the analysis. They help ensure clear delineation between prespecified and *ad hoc* analyses, and minimize opportunities for conducting unspecified analyses and selective reporting of results.

Individual-Level Confounder Summary Score-Based Methods

Confounder summary scores, such as propensity scores [86,87] and disease risk scores [88,89], are widely used in pharmacoepidemiologic research. If estimated correctly, these summary scores contain sufficient information to account for the confounding effects of the covariates used to estimate them. These data dimension reduction techniques have some appealing features useful for DDNs. Specifically, they obscure the information from a large number of covariates into scalar measures that are much less identifiable. Instead of requesting an individual-level dataset with information on individual covariates, one can replace these covariates with the summary scores [82,90,91]. In its simplest form, the dataset will only include variables indicating the treatment, outcome, follow-up (for time-to-event analysis), and confounder summary score. Other variables needed for the analysis, such as age or age categories if one wishes to perform age-stratified analysis, can also be requested. Conventional approaches to handling confounders, including matching, stratification, restriction, regression, and weighting, can then be done with the pooled, less granular, individual-level datasets.

This approach can perform essentially all the prespecified analyses afforded by the approach that shares individual confounder information, but it may not be able to accommodate all *ad hoc* analyses. For example, if sex is included in the estimation of the confounder summary score but is not requested separately, one will not be

able to perform a secondary, sex-stratified analysis without going back to the sites to request additional sex information.

The confounder summary scores should ideally be estimated and adjusted within each database, not just for practical reasons but also to ensure validity. For example, propensity scores are a function of the prevalence of the exposure in the population in which the scores are estimated. The prevalence of the exposure may vary across databases due to differences in formulary, regional prescribing pattern, and patient characteristics. Having site-specific propensity score models also allows the effect of a given covariate (e.g., age) on the probability of receiving the treatment of interest to vary by site. Researchers should account for site or database in the analysis by either including it as a stratification variable or performing within-site matching or stratification. As already discussed, data-adaptive approaches like the high-dimensional propensity score method [23] readily allow the analysis to be more tailored to the data availability at each participating site.

In contrast, the influence of risk factors on the outcome is generally more stable across databases, even if the outcome incidence varies by site. For example, the relation between age and heart failure, conditional on all other risk factors, should be similar across data sources. Therefore, it may be possible to combine disease risk scores, another commonly used confounder summary score that models and summarizes the associations between potential confounders and outcome risk, across sites. Additional research is needed to evaluate this issue.

Confounder Summary Score-Based Methods

It is possible to combine confounder summary score-based methods with other analytic techniques to further reduce the granularity of information shared. One can perform matching and stratification at the sites, and then only request the aggregate-level matched or stratified data to return to the analysis center [82,91]. In a matched analysis, if each site matches in the same fixed ratio, the only information needed for the analysis will be the total exposed and unexposed persons or person-times, and the number of exposed and unexposed outcomes. In a stratified analysis, participating sites send to the analysis center the total exposed and unexposed persons or person-times, and the number of exposed and unexposed outcomes within each stratum. Alternatively, one can structure the datasets into a risk set format at the sites and request risk set-based summary-level information for centralized analysis [82–84,92]. Results from the risk set-based approach have been shown to be statistically equivalent to results from the pooled individual-level stratified Cox regression model [83,92]. As with other methods, subgroup and sensitivity analyses will need to be prespecified so that appropriate summary-level information can be generated at the sites and shared for centralized analysis. As before, care should be taken with subgroup analyses to avoid small cells and potential identification risks. These methods protect against patient identification to an extent, but are not foolproof. If, for example, a rare disease or exposure is of interest, and prespecified analyses involve substantial stratification, some cells of the summary tables may be small enough to violate data partners' privacy regulations.

Meta-analysis of Database-Specific Results

An alternative to pooling individual-level data in a central repository is the commonly used approach of pooling site-specific effect estimates using meta-analytic techniques. In this approach, each site performs its own analysis, and the effect estimates and their variances (or other information needed to calculate database-specific weights) are provided to a central location and combined via meta-analysis [83,84,93–95]. The site-specific estimates can be obtained from matching, stratification, restriction, outcome modeling, or weighting, with or without confounder summary scores. This has

been shown to produce similar pooled effect estimates when compared with individual-level data analysis [90,96,97]. Although all data-sharing methods, including those discussed in this section, can in principle inspect treatment effect heterogeneity across databases, meta-analysis does that in the most obvious way, because database-specific effect estimates are shared and used in the pooled analysis. Each subgroup or sensitivity analysis requires all sites to perform each analysis internally, and then transfer the effect estimates to the lead team. Smaller sites may not be able to perform certain analyses, although sometimes using confounder summary scores to obtain site-specific effect estimate may help.

Distributed regression

The basic idea of distributed regression is for each data source to process its own individual-level data and share with the analysis center only summary statistics (e.g., sums of squares and cross-products matrix), such that the analysis center can either calculate the effect estimates or, if an iterative process is needed, update the parameter estimates and send them back to each data source to further update the summary statistics [98–101]. The iterative process continues until either a specified convergence criterion is met and the final parameter estimates are calculated, or the maximum number of iterations is reached. In other words, distributed regression conducts the same numeric algorithm with only centrally combined summary statistics as standard regression with pooled individual-level data. Although distributed regression is appealing in theory, it is relatively cumbersome to implement in practice, particularly for regression models that require multiple iterations. There are ongoing efforts to improve the practicality of distributed regression in existing DDNs [102–105].

Encryption

When applied in pharmacoepidemiology, encryption or hashing techniques are generally used to obscure potentially identifiable information while allowing valid database linkages [106,107]. In principle, it is possible to use these techniques to process the de-identified analytic dataset at the site before the encrypted data, along with the decryption method, are shared centrally for analysis. There have been some efforts in combining homomorphic encryption techniques with distributed regression [103,108], but using encryption to obscure potentially identifiable individual-level data is still quite theoretical and has not been widely implemented in practice.

The Future

More Sustainable and Efficient

DDNs, regardless of their actual configuration, typically require significant upfront investment. An architecture with a CDM is costly and time consuming to set up and maintain; however, efficiency benefits may be realized if the infrastructure is used for multiple studies. Some DDNs may achieve "economies of scale" and be able to conduct additional studies more efficiently and at a marginal cost compared to doing these studies as a series of "one-offs" [3,109]. DDNs without a CDM require infrastructure and initial investment to develop replicable and systematic processes. With more sustained funding support from regulatory agencies and other stakeholders, existing and future DDNs will have the much-needed stable foundation to grow, expand, and mature. This will be particularly important for some DDNs that are currently not supported by a single regulatory authority, such as AsPEN, and the various EU networks like PROTECT and EU-ADR. Even for funded DDNs, it can be argued that the infrastructure should be made available to other stakeholders with proper governance in place. The Sentinel system is an example of how this can be possible. Although the surveillance system was originally created by the FDA to meet its regulatory mandate, the agency envisioned the infrastructure eventually becoming a national resource for evidence generation [19]. Through the Innovation in Medical

Evidence Development and Surveillance initiative, non-FDA funders, including life science companies, can now access the same data sources and analytic tools used in Sentinel to conduct their own studies [110].

Broader Scope

With few exceptions, most DDNs were initially created for specific purposes (e.g., medical product safety surveillance). However, the scientific and technical infrastructure of many DDNs has the potential to address a wider range of topics, including comparative effectiveness research, patient-centered outcomes research, public health surveillance, and quality improvement. In addition to facilitating observational studies that analyze secondary data sources, some DDNs also support intervention studies. For example, PCORnet is conducting a pragmatic trial within participating health systems to examine aspirin dosing and secondary prevention of atherosclerotic cardiovascular disease [47]. Another pragmatic trial is underway in Sentinel to investigate the effect of direct mailings to patients and providers on initiation of anticoagulation therapy among eligible, treatment-naive patients [111]. These trials leverage the existing electronic healthcare databases of participating delivery systems or health plans to identify eligible patients and collect follow-up data, which allow the trials to be conducted more efficiently in real-world clinical settings compared to conventional randomized controlled trials.

More Diverse and Complementary Data Sources

Most DDNs are "horizontally partitioned," meaning that each database in the network contains information from different patients. Information is increasingly and routinely collected in various databases, for instance administrative claims databases, EHRs, disease or product registries, and data warehouses that contain information collected from wearables, mobile devices, or social media. In the future, we will likely see more DDNs that include disparate databases that include various data elements from the same individuals. These methods will require continued methodologic development to account for variation in data quality and completeness across data sources. Missing data and potential selection bias that arises from restricting the analysis to only patients appearing in multiple databases will require special attention.

More Robust and Secure Analysis

Continued methodologic advancement, both in developing cutting-edge analytic methods and in refining existing methods, will offer more analytic options that allow researchers to perform sophisticated statistical analysis while offering sufficient protection for patient confidentiality and data security. Existing methods already allow researchers to perform multivariable-adjusted outcome regression analysis and confounder summary-based analysis without sharing individual-level data for one-time exposures and one-time outcomes [82–84]. As DDNs mature, additional methodologic developments are likely to become available, including analysis of time-varying exposures, time-varying or repeated outcomes, missing data, and multilevel data.

Greater Transparency and Reproducibility

The DDN structure often requires researchers to prespecify the study design and analysis plan in advance, because they may not have direct access to all the data. For DDNs that employ a CDM, the analytic code will be pretested to ensure successful execution across the participating data partners. For DDNs that use a common protocol approach, the protocol will also need to be developed in advance. These data models, protocols, analytic code, and results should be made publicly available whenever

possible to improve transparency and encourage reproducibility [112]. Several DDNs, such as Sentinel and CNODES, are already adopting this policy, which has allowed other researchers to replicate the analyses [113,114]. OHDSI also makes its analytic tools publicly available.

Better Interoperability and Coordination across Networks

A possible future is one that has a national or international infrastructure that supports multiple DDNs. A healthcare delivery system or health plan can participate in multiple networks, each created for different purposes (e.g., medical product safety surveillance, comparative effectiveness research, pragmatic trials, public health surveillance). Each network will have its own governance and coordination. The networks can share the infrastructure, analytic tools, lessons learned, and software development and improvement. Regulators and decision-makers may also choose to collaborate on questions or work with multiple networks on specific queries.

References

1 AsPEN collaborators; Andersen M, Bergman U, Choi NK, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; **22**: 700–4.

2 Suissa S, Henry D, Caetano P, *et al*.; Canadian Network for Observational Drug Effects. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med* 2012; **6**: e134–40.

3 Steiner JF, Paolino AR, Thompson EE, Larson EB. Sustaining research networks: the twenty-year experience of the HMO research network. *EGEMS (Wash DC)* 2014; **2**: 1067.

4 Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014; **21**: 578–82.

5 Reynolds RF, Kurz X, de Groot MC, *et al*. The IMI PROTECT project: purpose, organizational structure, and procedures. *Pharmacoepidemiol Drug Saf* 2016; **25** (Suppl 1): 5–10.

6 Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative: a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther* 2016; **99**: 265–8.

7 McNeil MM, Gee J, Weintraub ES, *et al*. The Vaccine Safety Datalink: successes and challenges monitoring vaccine safety. *Vaccine* 2014; **32**: 5390–8.

8 Lai EC, Man KK, Chaiyakunapruk N, *et al*. Brief report: databases in the Asia-Pacific region: the potential for a distributed network approach. *Epidemiology* 2015; **26**: 815–20.

9 Pratt N, Andersen M, Bergman U, *et al*. Multi-country rapid adverse drug event assessment: the Asian Pharmacoepidemiology Network (AsPEN) antipsychotic and acute hyperglycaemia study. *Pharmacoepidemiol Drug Saf* 2013; **22**: 915–24.

10 Pratt N, Chan EW, Choi NK, *et al*. Prescription sequence symmetry analysis: assessing risk, temporality, and consistency for adverse drug reactions across datasets in five countries. *Pharmacoepidemiol Drug Saf* 2015; **24**: 858–64.

11 Roughead EE, Chan EW, Choi NK, *et al*. Variation in association between thiazolidinediones and heart failure across ethnic groups: retrospective analysis of large healthcare claims databases in six countries. *Drug Saf* 2015; **38**: 823–31.

12 Roughead EE, Chan EW, Choi NK, *et al*. Proton pump inhibitors and risk of *Clostridium difficile* infection: a multi-country study using sequence symmetry analysis. *Expert Opin Drug Saf* 2016; **15**: 1589–95.

13 Stang PE, Ryan PB, Racoosin JA, *et al*. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med* 2010; **153**: 600–6.

14 Chen RT, Glasser JW, Rhodes PH, *et al*.; Vaccine Safety Datalink Team. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. *Pediatrics* 1997; **99**: 765–73.

15 Chubak J, Ziebell R, Greenlee RT, *et al*. The Cancer Research Network: a platform for epidemiologic and health services research on cancer prevention, care, and outcomes in large, stable populations. *Cancer Causes Control* 2016; **27**: 1315–23.

16 Lu CY, Zhang F, Lakoma MD, *et al*. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: quasi-experimental study. *BMJ* 2014; **348**: g3596.

17 Go AS, Magid DJ, Wells B, *et al*. The Cardiovascular Research Network: a new paradigm for cardiovascular quality and outcomes research. *Circ Cardiovasc Qual Outcomes* 2008; **1**: 138–47.

18 Ross TR, Ng D, Brown JS, *et al*. The HMO research network virtual data warehouse: a public data model to support collaboration. *EGEMS (Wash DC)* 2014; **2**: 1049.

19 Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel system: a national resource for evidence development. *N Engl J Med* 2011; **364**: 498–9.

20 Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel network: improving the evidence of medical-product safety. *N Engl J Med* 2009; **361**: 645–7.

21 Platt R, Carnahan RM, Brown JS, *et al*. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 1–8.

22 Gagne JJ, Han X, Hennessy S, *et al*. Successful comparison of US Food and Drug Administration Sentinel analysis tools to traditional approaches in quantifying a known drug-adverse event association. *Clin Pharmacol Ther* 2016; **100**: 558–64.

23 Zhou M, Wang SV, Leonard CE, *et al*. Sentinel modular program for propensity score-matched cohort analyses: application to glyburide, glipizide, and serious hypoglycemia. *Epidemiology* 2017; **28**: 838–46.

24 Connolly JG, Wang SV, Fuller CC, *et al*. Development and application of two semi-automated tools for targeted medical product surveillance in a distributed data network. *Curr Epidemiol Rep* 2017; **4**: 298–306.

25 Gagne JJ, Houstoun M, Reichman ME, Hampp C, Marshall JH, Toh S. Safety assessment of niacin in the US Food and Drug Administration's Mini-Sentinel system. *Pharmacoepidemiol Drug Saf* 2018; **27**: 30–7.

26 Davis RL, Kolczak M, Lewis E, *et al*. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005; **16**: 336–41.

27 Yih WK, Kulldorff M, Fireman BH, *et al*. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink project. *Pediatrics* 2011; **127**(Suppl 1): S54–S64.

28 Fahey KR. The pioneering role of the Vaccine Safety Datalink Project (VSD) to advance collaborative research and distributed data networks. *EGEMS (Wash DC)* 2015; **3**: 1195.

29 Lieu TA, Kulldorff M, Davis RL, *et al*. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45**: S89–S95.

30 Greene SK, Kulldorff M, Yin R, *et al*. Near real-time vaccine safety surveillance with

partially accrued data. *Pharmacoepidemiol Drug Saf* 2011; **20**: 583–90.
31 Madigan D, Ryan PB, Schuemie M, *et al*. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol* 2013; **178**: 645–51.
32 Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med* 2012; **31**: 4401–15.
33 Hripcsak G, Duke JD, Shah NH, *et al*. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; **216**: 574–8.
34 Coloma PM, Schuemie MJ, Trifiro G, *et al*. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR project. *Pharmacoepidemiol Drug Saf* 2011; **20**: 1–11.
35 Arfe A, Scotti L, Varas-Lorenzo C; Safety of Non-steroidal Anti-inflammatory Drugs Project. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case-control study. *BMJ* 2016; **354**: i4857.
36 CORDIS. Arrhythmogenic Potential of Drugs (ARITMO): project information. 2017. https://cordis.europa.eu/project/rcn/94061/factsheet/en (accessed May 2019).
37 CORDIS. Safety Evaluation of Adverse Reactions in Diabetes (SAFEGUARD): project information. https://cordis.europa.eu/project/rcn/100121/reporting/en (accessed May 2019).
38 CORDIS. Global Research in Paediatrics (GRIP): project information. https://cordis.europa.eu/project/rcn/97619/factsheet/en (accessed May 2019).
39 Accelerated Development of Vaccine Benefit-risk Collaboration in Europe (ADVANCE). About ADVANCE. http://www.advance-vaccines.eu (accessed April 2018).
40 Shin JY, Roughead EE, Park BJ, Pratt NL. Cardiovascular safety of methylphenidate among children and young people with attention-deficit/hyperactivity disorder (ADHD): nationwide self controlled case series study. *BMJ* 2016; **353**: i2550.
41 Dormuth CR, Hemmelgarn BR, Paterson JM, *et al*.; Canadian Network for Observational Drug Effects. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases. *BMJ* 2013; **346**: f880.
42 Filion KB, Azoulay L, Platt RW, *et al*. A multicenter observational study of incretin-based drugs and heart failure. *N Engl J Med* 2016; **374**: 1145–54.
43 Henry D, Dormuth C, Winquist B, *et al*. Occurrence of pregnancy and pregnancy outcomes during isotretinoin therapy. *CMAJ* 2016; **188**: 723–30.
44 Graham DJ, Staffa JA, Shatin D, *et al*. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *JAMA* 2004; **292**: 2585–90.
45 Davis RL, Rubanowice D, McPhillips H, *et al*. Risks of congenital malformations and perinatal events among infants exposed to antidepressant medications during pregnancy. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1086–94.
46 Habel LA, Cooper WO, Sox CM, *et al*. ADHD medications and risk of serious cardiovascular events in young and middle-aged adults. *JAMA* 2011; **306**: 2673–83.
47 Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: shining light on a new research paradigm. *Ann Intern Med* 2015; **163**: 635–6.
48 Block JP, Bailey LC, Gillman MW, *et al*.; Antibiotics and Childhood Growth Study. The PCORnet Antibiotics and Childhood Growth Study: process for cohort creation and cohort description. *Acad Pediatr* 2018; **18**(5): 569–76.

49 Toh S, Rasmussen-Torvik LJ, Harmata EE, *et al.*; Collaborative PCBS. The National Patient-Centered Clinical Research Network (PCORnet) bariatric study cohort: rationale, methods, and baseline characteristics. *JMIR Res Protoc* 2017; **6**: e222.

50 Brauer R, Douglas I, Garcia Rodriguez LA, *et al.* Risk of acute liver injury associated with use of antibiotics: comparative cohort and nested case-control studies using two primary care databases in Europe. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 29–38.

51 Souverein PC, Abbing-Karahagopian V, Martin E, *et al.* Understanding inconsistency in the results from observational pharmacoepidemiological studies: the case of antidepressant use and risk of hip/femur fractures. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 88–102.

52 Schuerch M, Gasse C, Robinson NJ, *et al.* Impact of varying outcomes and definitions of suicidality on the associations of antiepileptic drugs and suicidality: comparisons from UK Clinical Practice Research Datalink (CPRD) and Danish National Registries (DNR). *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1):142–55.

53 Toh S, Reichman ME, Graham DJ, *et al.*; Mini-Sentinel Saxagliptin AMISWG. Prospective postmarketing surveillance of acute myocardial infarction in new users of saxagliptin: a population-based study. *Diabetes Care* 2018; **41**: 39–48.

54 Go AS, Singer DE, Toh S, *et al.* Outcomes of dabigatran and warfarin for atrial fibrillation in contemporary practice: a retrospective cohort study. *Ann Intern Med* 2017; **167**: 845–54.

55 Yih WK, Lieu TA, Kulldorff M, *et al.* Intussusception risk after rotavirus vaccination in U.S. infants. *N Engl J Med* 2014; **370**: 503–12.

56 Thompson WW, Price C, Goodson B, *et al.* Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years. *N Engl J Med* 2007; **357**: 1281–92.

57 Lee GM, Greene SK, Weintraub ES, *et al.*; Vaccine Safety Datalink Project. H1N1 and seasonal influenza vaccine safety in the Vaccine Safety Datalink project. *Am J Prev Med* 2011; **41**: 121–8.

58 Gee J, Sukumaran L, Weintraub E; Vaccine Safety Datalink Project. Risk of Guillain-Barré syndrome following quadrivalent human papillomavirus vaccine in the Vaccine Safety Datalink. *Vaccine* 2017; **35**: 5756–8.

59 Maro JC, Brown JS, Dal Pan GJ, Li L. Orphan therapies: making best use of postmarket data. *J Gen Intern Med* 2014; **29**(Suppl 3): S745–51.

60 Kesselheim AS, Gagne JJ. Strategies for postmarketing surveillance of drugs for rare diseases. *Clin Pharmacol Ther* 2014; **95**: 265–8.

61 Kothari C, Wack M, Hassen-Khodja C, *et al.* Phelan-McDermid syndrome data network: integrating patient reported outcomes with clinical notes and curated genetic reports. *Am J Med Genet B Neuropsychiatr Genet* 2018; **177**(7): 613–24.

62 Sejvar JJ, Baughman AL, Wise M, Morgan OW. Population incidence of Guillain-Barré syndrome: a systematic review and meta-analysis. *Neuroepidemiology* 2011; **36**: 123–33.

63 Toh S, Hampp C, Reichman ME, *et al.* Risk for hospitalized heart failure among new users of saxagliptin, sitagliptin, and other antihyperglycemic drugs: a retrospective cohort study. *Ann Intern Med* 2016; **164**: 705–14.

64 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Characterising the risk of major bleeding in patients with non-valvular atrial fibrillation: non-interventional study of patients taking direct oral anticoagulants in the EU. http://www.encepp.eu/encepp/viewResource.htm?id=21042 (accessed April 2018).

65 DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med* 1994; **13**: 1341–52; discussion 53–6.

66 Lan KKG, DeMets DL. Discrete sequential boundries for clinical trials. *Biometrika* 1983; **70**: 639–63.

67 Lan KK, DeMets DL. Changing frequency of interim analysis in sequential monitoring. *Biometrics* 1989; **45**: 1017–20.

68 Chrischilles EA, Gagne JJ, Fireman B, *et al*. Prospective surveillance pilot of rivaroxaban safety within the US Food and Drug Administration Sentinel System. *Pharmacoepidemiol Drug Saf* 2018; **27**: 263–71.

69 Mazor KM, Richards A, Gallagher M, *et al*. Stakeholders' views on data sharing in multicenter studies. *J Comp Eff Res* 2017; **6**(6): 537–47.

70 Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care* 2010; **48**: S45–S51.

71 Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff (Millwood)* 2009; **28**: 454–66.

72 Curtis LH, Weiner MG, Boudreau DM, *et al*. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 23–31.

73 PCORnet. PCORnet Common Data Model (CDM). 2018. http://www.pcornet.org/pcornet-common-data-model (accessed April 2018).

74 Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012; **19**: 54–60.

75 Bell C, Chakravarty A, Gruber S, *et al*. Characteristics of study design and elements that may contribute to the success of electronic safety monitoring systems. *Pharmacoepidemiol Drug Saf* 2014; **23**: 1223–5.

76 Gruber S, Chakravarty A, Heckbert SR, *et al*. Design and analysis choices for safety surveillance evaluations need to be tuned to the specifics of the hypothesized drug-outcome association. *Pharmacoepidemiol Drug Saf* 2016; **25**: 973–81.

77 Xu Y, Zhou X, Suehs BT, *et al*. A comparative assessment of Observational Medical Outcomes Partnership and Mini-Sentinel common data models and analytics: implications for active drug safety surveillance. *Drug Saf* 2015; **38**: 749–65.

78 Gagne JJ. Common models, different approaches. *Drug Saf* 2015; **38**: 683–6.

79 Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*, 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.

80 Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006; **60**: 578–86.

81 Rassen JA, Moran J, Toh D, *et al*. Evaluating strategies for data sharing and analyses in distributed data settings. 2013. https://www.sentinelinitiative.org/sites/default/files/Methods/Mini-Sentinel_Methods_Evaluating-Strategies-for-Data-Sharing-and-Analyses_0.pdf (accessed April 2018).

82 Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med Care* 2013; **51**: S4–S10.

83 Toh S, Shetterly S, Powers JD, Arterburn D. Privacy-preserving analytic methods for multisite comparative effectiveness and patient-centered outcomes research. *Med Care* 2014; **52**: 664–8.

84 Toh S, Reichman ME, Houstoun M, *et al*. Multivariable confounding adjustment in distributed data networks without sharing of patient-level data. *Pharmacoepidemiol Drug Saf* 2013; **22**: 1171–7.

85 Toh S, Platt R. Is size the next big thing in epidemiology? *Epidemiology* 2013; **24**: 349–51.

86 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.

87 Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984; **79**: 516–24.

88 Hansen BB. The prognostic analogue of the propensity score. *Biometrika* 2008; **95**: 481–8.

89 Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res* 2009; **18**: 67–80.

90 Rassen JA, Avorn J, Schneeweiss S. Multivariate-adjusted pharmacoepidemiologic analyses of confidential information pooled from multiple health care utilization databases. *Pharmacoepidemiol Drug Saf* 2010; **19**: 848–57.

91 Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010; **19**: 858–68.

92 Fireman B, Lee J, Lewis N, Bembom O, van der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. *Am J Epidemiol* 2009; **170**: 650–6.

93 Platt RW, Dormuth CR, Chateau D, Filion K. Observational studies of drug safety in multi-database studies: methodological challenges and opportunities. *EGEMS (Wash DC)* 2016; **4**: 1221.

94 Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 501, updated September 2008. London: Cochrane Collaboration, 2008, Chapter 9.

95 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–88.

96 Toh S, Reichman ME, Houstoun M, *et al*. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. *Arch Intern Med* 2012; **172**: 1582–9.

97 Li X, Fireman BH, Curtis JR, *et al*. Validity of privacy-protecting analytical methods that use only aggregate-level information to conduct multivariable-adjusted analysis in distributed data networks. *Am J Epidemiol* 2019; **188**(4): 709–23.

98 Karr AF, Lin X, Sanil AP, Reiter JP. Secure regression on distributed databases. *J Comput Graph Stat* 2005; **14**: 263–79.

99 Karr AF, Fulp WJ, Vera F, Young SS, Lin X, Reiter JP. Secure, privacy-preserving analysis of distributed databases. *Technometrics* 2007; **49**: 335–45.

100 Fienberg SE, Fulp WJ, Slavković AB, Wrobel TA. "Secure" log-linear and logistic regression analysis of distributed databases. *Lect Notes Comput Sci* 2006; **2006**: 277–90.

101 Slavković AB, Nardi Y, Tibbits MM. Secure logistic regression of horizontally and vertically partitioned distributed databases. *Proceedings of Workshop on Privacy and Security Aspects of Data Mining*. Washington, DC: IEEE Computer Society Press, 2007, pp. 723–8.

102 Meeker D, Jiang X, Matheny ME, *et al*. A system to build distributed multivariate models and manage disparate data sharing policies: implementation in the scalable national network for effectiveness research. *J Am Med Inform Assoc* 2015; **22**: 1187–95.

103 El Emam K, Samet S, Arbuckle L, Tamblyn R, Earle C, Kantarcioglu M. A secure distributed logistic regression protocol for the detection of rare adverse drug events. *J Am Med Inform Assoc* 2012; **20**: 453–61.

104 Wolfson M, Wallace SE, Masca N, *et al*. DataSHIELD: resolving a conflict in contemporary bioscience – performing a pooled analysis of individual-level data without sharing the data. *Int J Epidemiol* 2010; **39**: 1372–82.

105 Her QL, Malenfant JM, Malek S, *et al*. A query workflow design to perform automatable distributed regression analysis in large distributed data networks. *EGEMS (Wash DC)* 2018; **6**(1): 11.

106 Kho AN, Cashy JP, Jackson KL, *et al*. Design and implementation of a privacy preserving electronic health record linkage tool in

Chicago. *J Am Med Inform Assoc* 2015; **22**: 1072–80.

107 Bezin J, Duong M, Lassalle R, *et al*. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017; **26**: 954–62.

108 Hall R, Fienberg SE, Nardi Y. Secure multiple linear regression based on homomorphic encryption. *J Off Stat* 2011; **27**: 669–91.

109 Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood)* 2014; **33**: 1178–86.

110 Reagan-Udall Foundation for the Food and Drug Administration. Innovation in medical evidence development and surveillance. http://reaganudall.org/innovation-medical-evidence-development-and-surveillance (accessed April 2018).

111 Pokorney SD, Cocoros N, Al-Khalidi H, *et al*. Implementation of a randomized controlled trial to improve treatment with oral anticoagulants in patients with atrial fibrillation (IMPACT-AFib). 2017. https://www.sentinelinitiative.org/sites/default/files/IMPACT-AFib_Protocol_v3.pdf (accessed April 2018).

112 Wang SV, Schneeweiss S, Berger ML, *et al*., Joint I-ISTFoRWEiHCDM. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf* 2017; **26**: 1018–32.

113 Li X, Girman CJ, Ofner S, *et al*. Sensitivity analysis of methods for active surveillance of acute myocardial infarction using electronic databases. *Epidemiology* 2015; **26**: 130–2.

114 Simeone JC, Nordstrom BL, Appenteng K, Huse S, D'Silva M. Replication of Mini-Sentinel study assessing mirabegron and cardiovascular risk in non-Mini-Sentinel databases. *Drugs Real World Outcomes* 2018; **5**: 25–34.

26

Comparative Effectiveness Research

Soko Setoguchi[1] *and Ian Chi Kei Wong*[2,3]

[1] *Rutgers Robert Wood Johnson Medical School, Rutgers Center for Pharmacoepidemiology and Treatment Science, New Brunswick, NJ, USA*
[2] *Department of Pharmacology and Pharmacy, University of Hong Kong, Hong Kong*
[3] *UCL School of Pharmacy, London, UK*

Comparative Effectiveness Research in the US

Clinical and scientific communities have a long-standing desire to use scientific evidence to optimize clinical decisions for patients. While randomized controlled trials (RCTs) and meta-analyses of RCTs are generally considered to constitute the highest level of evidence, they have also been criticized for several aspects inherent to their design: comparison to placebo rather than alternative treatment; nonrepresentativeness of patient populations that tend to exclude older and multimorbid adults and children; controlled settings that differ from real-world care settings; relatively short follow-up; frequent use of surrogate endpoints rather than hard endpoints such as clinical events or death; and insufficient sample size to assess subgroup effects. Efforts to produce evidence that overcomes these limitations and is directly applicable to real-world patients as well as making clinical care more rational have been referred to at different times as *outcomes research*, *effectiveness research*, *evidence-based research*, *health technology assessment*, and, most recently, *comparative effectiveness research* (CER) [1]. To reduce the perception that the main agenda behind the push for CER is cost containment for healthcare, at least one government agency has begun relabeling CER as *patient-centered health research* [2].

CER is not a new concept, but has existed for the last several decades under various labels, and its popularity in the US had risen in response to several government initiatives. Earlier government initiatives for CER in the US were attempted first by the Congressional Office of Technology Assessment (established in 1972), then by the National Center for Health Care Technology (1978–1982), and then by the Agency for Health Care Policy and Research (established in 1989 and later renamed the Agency for Healthcare Research and Quality, AHRQ) [3]. The most recent impetus for CER came from the 2009 American Recovery and Reinvestment Act (ARRA Stimulus), with an appropriation of $1.1 billion "to study the comparative effectiveness of healthcare treatments" [4].

Furthermore, in 2010, the Patient Protection and Affordable Care Act (PPACA) authorized

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

the establishment of the Patient-Centered Outcomes Research Institute (PCORI) to carry out CER and improve its quality and relevance [5]. PPACA established new requirements for the Department of Health and Human Services (HHS) to disseminate findings from federally funded CER, including findings published by PCORI, and to coordinate with relevant federal health programs to build data capacity for this research. To fund CER activities, PPACA established the Patient-Centered Outcomes Research Trust Fund (PCORTF), from which PCORI and HHS are expected to receive an estimated $4 billion from fiscal years 2010 through 2019. As of November 2017, PCORI had disbursed more than $2 billion for approximately 600 CER-related projects [6].

CER in Europe and Other Countries

Europe

In recent years a number of European Union (EU) countries have introduced so-called health technology assessments (HTA). HTA includes not only assessment of clinical effectiveness, but cost-effectiveness as well [7]. Publicly funded healthcare systems are the main healthcare providers in a number of EU countries, and these systems are under substantial financial pressure to make the best use of available resources. Assessing cost-effectiveness as part of HTA is therefore critical in the evaluation of health technology.

The National Institute for Health and Care Excellence (NICE) in England and Wales, created in 1999, represents one model for using CER primarily to inform policy and practice, but also to develop research recommendations [8]. Since April 2013, NICE has gained new responsibilities for providing guidance to those working in social care. Accordingly, NICE guidance documents are used by the National Health Service (NHS), local government, employers, volunteer groups, and others involved in delivering care or promoting wellbeing [9]. NICE guidance takes several forms, including NICE guidelines, technology appraisals guidance, medical technologies, and diagnostics guidance, as described shortly.

NICE guidelines make evidence-based recommendations on a wide range of topics, from preventing and managing specific conditions, improving health, and managing medicines in different settings, to providing social care to adults and children and planning broader services and interventions to improve the health of communities. These guidelines aim to promote integrated care where appropriate. NICE has provided a substantial number of evidence-based guidelines for clinical practice [10], though not without controversy and challenge [11].

Technology appraisals guidance assesses the clinical and cost-effectiveness of health technologies, such as new pharmaceutical and biopharmaceutical products, but also procedures, devices, and diagnostic agents. For example, recent guidance recommends ixazomib with lenalidomide and dexamethasone for use within the Cancer Drugs Fund (a central funding source for cancer drugs in England) as an option for treating multiple myeloma in adults only if patients have already had two or three lines of therapy [12]. Technology appraisals guidance is intended to ensure that all NHS patients have equitable access to the most clinically and cost-effective treatments that are viable.

Medical technologies and diagnostics guidance helps to ensure that the NHS is able to adopt clinically and cost-effective technologies rapidly and consistently. For example, Neuropad is a technology that aims to detect preclinical diabetic peripheral neuropathy. However, its use is not supported by evidence [13]. Interventional procedures guidance provides recommendations on whether interventional procedures are effective and safe enough for use in the NHS. For example, NICE recently

recommended that intravesical microwave hyperthermia and chemotherapy for non-muscle-invasive bladder cancer be used only with special arrangements for clinical governance, consent, and audit or research [14].

For their evaluations, NICE's advisory committees use objective evidence provided by academic institutions in the UK, such as the Royal College of Physicians, under contract with NICE to perform evidence syntheses and to conduct small-scale studies entailing primary data collection [8]. The explicit use of cost-effectiveness data to evaluate and choose among medical interventions is viewed in the UK "as a tool to ensure fair shares for all in a resource-limited system," according to Chalkidou and Walley [8].

A six-country comparison by Sorenson [15] and a similar three-country comparison by Evans [16] illustrate the considerable efforts extended by European governments to incorporate CER into health policy decisions and the different approaches used for organizing these efforts. In France (the National Authority for Health – Haute Autorité de Santé or HAS [17]), Germany (the Institute for Quality and Efficiency in Healthcare – Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen or IQWiG [18]), and the Netherlands (Commissie Farmaceutische Hulp or CHF, Committee for Pharmaceutical Aid), the entities responsible for CER, act in an advisory role to the government, making recommendations on reimbursements and pricing. This is in contrast with the UK (NICE), Denmark (Reimbursement Committee of the Danish Medicines Agency or DKMA), and Sweden (Dental and Pharmaceutical Benefits Board or TLV), where the CER entities have regulatory authority and are directly responsible for prioritizing reimbursements for drug and devices [15,16]. Cost-effectiveness data are formally incorporated in evaluations and recommendations about coverage and pricing by most CER entities (UK, Germany, the Netherlands, and Sweden) [15].

Another five-country comparison by Levy *et al.* [19] that also includes Canada and Australia (the Pharmaceutical Benefits Advisory Committee, or PBAS [20]) noted that in each of the countries surveyed, the health technology evaluation committees (conceptually comparable to CER) retain their independence regarding decisions about which technologies are included in the formulary, despite receiving government funding. Members of these committees are primarily health professionals, with only Canada, Australia, and Scotland also including public representatives, and only Scotland permitting industry representation as well [19].

Other Countries

Healthcare systems and their financing mechanisms outside Western Europe and North America are very diverse, and it is impossible to comprehensively discuss the applications (or potential applications) of CER in all countries. According to Bloomberg Health Care Efficiency, Hong Kong was ranked the most efficient healthcare system in the world in 2017 and 2018 [21]. Hong Kong has a universal, publicly funded healthcare system which does not formally apply CER in decision-making. In parallel, Hong Kong also has a very well-developed private healthcare system, funded by insurance and patient out-of-pocket payments. The two sectors complement each other, in that the private sector is the major provider of primary healthcare, while the public sector is the predominant provider of secondary and tertiary healthcare services. About 70% of outpatient consultations are provided by the private sector, while over 90% of inpatient services (in terms of the number of bed-days) are provided by public hospitals. This system presents substantial difficulties for effectiveness research, as the two sectors cannot use the same CER for evaluation. At present, neither the public nor the private sector has adopted formal assessment of CER for evaluation of treatment. This raises important

questions regarding what factors beyond CER play significant roles in the efficiency of the healthcare system.

African countries in general are facing significant issues in healthcare financing and are struggling to provide sufficient publicly funded healthcare services. In South Africa, the National Department of Health includes a series of explicit references to HTA in a white paper setting out the government's 10-year vision for high-quality universal healthcare coverage. A dedicated taskforce has been set up to consider HTA and other tools in order to design high-quality, affordable packages of health services [22].

Patel *et al.* describe the healthcare and government environment and the use (and potential use) of CER to control healthcare expenditures in China, India, and South Korea [23]. This report demonstrates the diversity of the healthcare systems and potential uses of CER in these three countries. CER will clearly be of increasing importance to aid government agencies in healthcare resource allocation. While the use of CER by government agencies has been well established for a substantial period outside the US, much of the recent activity is occurring within the US, and that will be the primary focus of this chapter.

Efficacy vs. Effectiveness

A study of *treatment efficacy* investigates whether a drug *has the ability to* bring about a given intended effect in ideal (controlled) settings. For example, a drug efficacy study would be centered on the question: "In an ideal world, with perfect adherence, no interactions with other drugs or other diseases, etc., *could* the drug achieve its intended effects?" In contrast, a study of *treatment effectiveness* investigates whether, in real-world patients and settings, a treatment *in fact* achieves its desired effect. For example, a drug given in a controlled setting may be shown to reduce glucose levels in younger patients having no major co-morbidities, but it might not achieve good glucose control in older patients with heart failure if it causes even mild water retention that leads to nonadherence or premature discontinuation. To answer questions about effectiveness, studies need to include representative real-world patients and assess effectiveness in real-world care settings.

Definitions, Key Components, and Goals of CER

CER seeks to assist stakeholders, for example patients, clinicians, insurers, the medical products industry, and policymakers to make informed decisions to improve healthcare at both individual and population levels. Several definitions of CER have been proposed by US government and nongovernment organizations and are summarized in Table 26.1. In Europe, the term CER is not commonly used, but HTA describes similar though not identical research. Several definitions of HTA are also provided in Table 26.1. HTA as defined by the UK National Institute for Health Research (NIHR) is actually broader than CER, since it formally includes cost-effectiveness evaluation, whereas CER generally does not.

For CER to assist in clinical decision-making, it must include three key components: (i) evidence synthesis (identifying and summarizing already existing data addressing a question); (ii) evidence generation (creating new data addressing a question); and (iii) evidence dissemination (distributing the available data with the goal of informing healthcare decision-making). In other words, for some decisions, existing evidence from individual studies may be controversial or insufficient to support specific clinical decisions. In such cases, the evidence must be synthesized (evidence synthesis), which may then provide a sufficient basis to support the decision or identify knowledge gaps to guide

Table 26.1 Definitions of comparative effectiveness research and health technology assessment proposed by US and other government and nongovernment organizations.

Agency/report	Definition
Comparative effectiveness research (CER)	
US Congressional Budget Office report, December 2007 [24]	"A rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients. Such a study may compare similar treatments, such as competing drugs, or it may analyze very different approaches, such as surgery and drug therapy."
Institute of Medicine report [25]	"The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist patients, clinicians, purchasers, policy makers, and the public to make informed decisions that will improve health care at both the individual and population levels."
US Federal Coordinating Council [26]	"CER is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in 'real world' settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances."
Patient-Centered Outcomes Research Institute [27]	"CER is a field of research designed to compare the effectiveness of two or more interventions or approaches to health care, examining their risks and benefits. CER findings assist clinicians, patients and other stakeholders in making informed decisions that improve health care for both individuals and populations. The direct comparison of two or more interventions distinguishes CER from studies explor[ing] outcomes related to one intervention alone. CER can not only validate a particular intervention but also identify which of available treatments best meet the needs of a given population."
Health technology assessment (HTA)	
National Institute for Health Research, UK [7]	"HTA research is undertaken when evidence exists to show that a technology can be effective. The purpose of an HTA study is to establish the clinical and cost-effectiveness for the NHS in comparison with the current best alternative(s). A study may also investigate uncertainty around a technology's place in the existing care pathway. 'Technologies' in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease."
European Commission [28]	"HTA measures the added value of a new health technology compared to existing ones. Examples of health technologies include medicinal products, medical equipment, diagnostic and treatment methods, rehabilitation, and prevention methods."
International Network of Agencies for Health Technology Assessment (INAHTA) [29]	"HTA is the systematic evaluation of the properties and effects of a health technology, addressing the direct and intended effects of this technology, as well as its indirect and unintended consequences, and aimed mainly at informing decision making regarding health technologies. HTA is conducted by interdisciplinary groups that use explicit analytical frameworks drawing on a variety of methods."

further evidence generation. For some decisions faced by patients, clinicians, insurers, and policymakers, there may be insufficient evidence from individual studies to inform the decision. In these cases, new CER studies must be conducted to generate evidence (evidence generation). Generated or synthesized evidence must be disseminated for decision-makers of healthcare and CER to make informed decisions. It is also important to note that CER may assess both benefit and harms. Therefore, traditional pharmacoepidemiologic studies assessing the safety (harms) of medications in postmarket settings fall under the umbrella of CER.

The most important gaps in the current knowledge base about treatment interventions are lack of information about how a treatment works in actual clinical practice in contrast to the artificial settings of clinical trials, lack of information about the comparative effectiveness of treatment options, and lack of information about how variation in patient characteristics affects treatment effectiveness [30]. CER has the potential to fill important evidence gaps associated with the limitations of a predominantly RCT-driven drug and device approval pathway. The RCT pathway speaks to efficacy rather than effectiveness, because (i) placebo is often used rather than an active comparator agent; (ii) RCT study populations are not representative of the medication users post-approval (i.e., RCTs tend to exclude older and multimorbid adults and children); and (iii) the controlled settings used in clinical development programs often differ substantially from real-world care settings (e.g., they use short follow-up and surrogate endpoints).

In summary, the goals of CER are (i) to inform decisions on interventions or approaches to health care in real-world settings with regard to their intended and unintended outcomes that are relevant to patients; (ii) to put new technology into proper perspective in relation to older technology; and (iii) to identify patients who are more or less likely to respond to some interventions than others [31]. As a result, CER is expected to increase the use of more effective clinical options and decrease the use of less effective treatments [1,32–34]. Another consequence of achieving these goals could be a reduction in healthcare costs through avoidance of treatments that do not work or are less effective than alternatives.

CER and Pharmacoepidemiology

The concept of CER is in fact very familiar to pharmacoepidemiologists. Classic pharmacoepidemiologic studies that assess postmarket safety of medications constitute CER as defined earlier. Also, soon after the field of pharmacoepidemiology emerged in response to the need to study drug safety after marketing of medications, pharmacoepidemiologists recognized the need for postmarketing "efficacy" assessment (now defined as "effectiveness") and debated the challenges of assessing "intended" effects or benefits [35–37]. Despite the concern that non-experimental studies may not be useful in studying the intended effects of drugs, Strom *et al.* showed that of the 100 most recently approved drugs with 131 potential drug uses, only 28% would require experimental designs [38]. In the field of pharmacoepidemiology, we developed a research framework for experimental and non-experimental studies, knowledge of study designs, data sources, and analytic strategies, and faced various new methodologic challenges when studying unintended and intended effects in real-world patients. As described in the previous section, CER became a popular concept and a well-funded field as a result of the most recent government initiative in 2008 and the subsequent establishment of PCORI. The majority of CER has dealt with the effectiveness of medications, surgical procedures, and

medical devices, which is another reason why the field is of great relevance and interest to pharmacoepidemiologists.

In the context of pharmacoepidemiology and especially in the drug development process, CER covers the tail end of the pathway that begins with bench research (characterized by preclinical research to qualify for Phase I regulatory approval), moving to bedside research (characterized by proof of concept and efficacy research to qualify for Phase II regulatory approval), and ending with population research (characterized by clinical efficacy to qualify for Phase III regulatory approval and Phase IV clinical safety and effectiveness in postmarket settings), and finally research on the effect of policies (characterized by postmarketing surveillance and pharmacoeconomic research). A schematic illustration of this process is presented in Figure 26.1. In one sense, the full scope of CER is much broader than pharmacoepidemiology, as CER covers a range of clinical modalities for prevention, diagnosis, and treatment (drugs, medical devices, procedures, behavioral and other complex social interventions, as well as health delivery systems and policies) [39]. It also covers strategies for implementation. In another sense, however, CER is narrower than pharmacoepidemiology, not only because it covers the tail end of the pharmacoepidemiology spectrum (Figure 26.1), but also because it emphasizes "head-to-head" comparisons of the safety and benefits of treatments and diagnostic strategies to identify "best-in-class" treatments in the real world [40], whereas pharmacoepidemiology can compare users to nonusers or to alternative treatments.

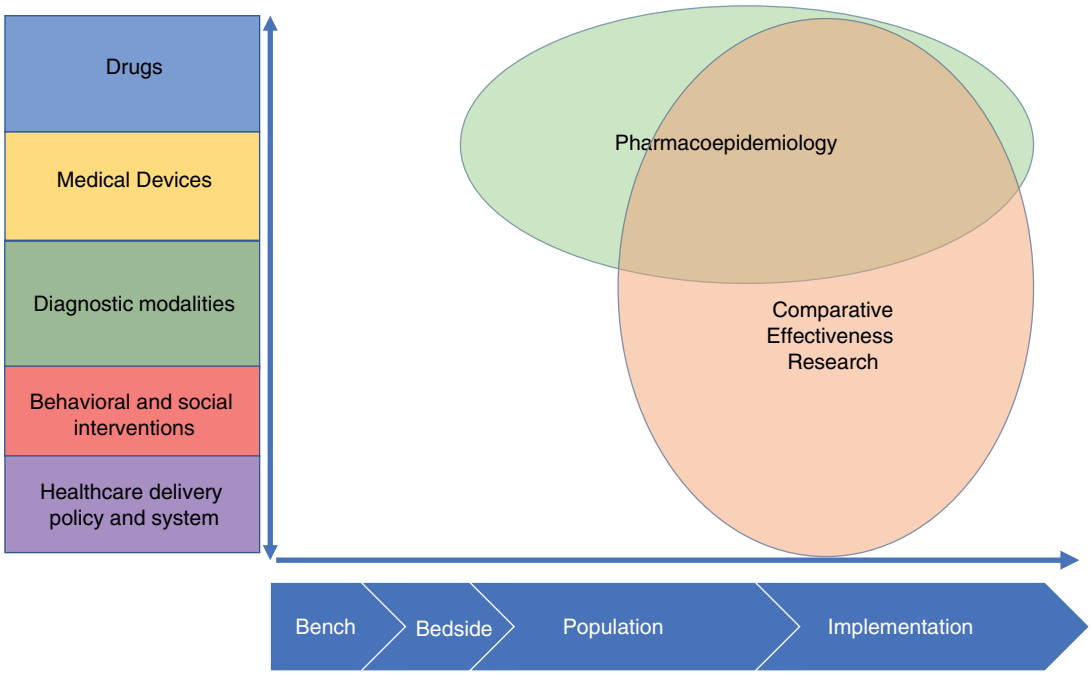


Figure 26.1 Schematic representation of overlap between pharmacoepidemiology and comparative effectiveness research.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Scope of CER

CER is broad in scope and addresses the continuum of medical and surgical interventions, including drugs, biologics, devices, medical procedures, technologies, behavioral interventions, prevention strategies, talk therapies, diagnostics, complex social interventions, and health delivery systems [25,41]. In addition, as characterized by Lauer and Collins [42], CER should utilize an array of technologies that enable quality and efficient healthcare delivery, and should account for the wide range of infrastructure of integrated healthcare systems. CER also encompasses beneficial and adverse effects as well as economic implications. It focuses attention not only on knowledge creation, but also on strategies for implementation. This broad scope can only be addressed by a diverse research portfolio that employs multiple study designs and analytic techniques (randomized trials, observational studies, and meta-analyses), as well as diverse data from primary data collection, preexisting data, and hybrid approaches linking different data sources [32,42].

Key Attributes of CER

The key attributes of CER that are embedded explicitly or implicitly in the aforementioned definitions and goals are as follows:

- It studies effectiveness in real-world patients and settings.
- It directly compares alternative methods to prevent, diagnose, treat, and monitor clinical conditions (rarely comparing alternatives to placebo, as "doing nothing" is often not a real-world clinical decision).
- It involves stakeholders, including patients and caregivers, in the research process.
- It uses clinically relevant and patient-centered outcomes.
- It assesses subgroups and different care settings in which differential effects may be observed (so that the evidence is more applicable to individual patients and is useful in various clinical settings).

Key attributes and related goals of CER studies are presented in Table 26.2.

The first attribute (inclusion of real-world patients and settings) is necessary to increase the direct applicability of the evidence generated from CER. Traditional efficacy trials are typically conducted by investigators affiliated with tertiary care hospitals. In contrast, CER should include data from patients and physicians from a wide range of care settings. The vision for CER is that it will provide opportunities for community hospitals and practices to become involved [43].

In real-world clinical practice, clinicians and patients need information to understand the comparative benefit or safety of two or more alternative treatments and to choose the best option. Therefore, "doing nothing" (placebo) is infrequently a viable alternative to treatment. However, traditional efficacy trials compare an intervention to nonintervention, for example treatment A to placebo, and are thus not informative on the comparative effect of different treatments. The goal of CER, as already mentioned, is to inform clinical or policy decisions *among alternative options*. Therefore, head-to-head comparison of alternative methods (including nonintervention if that is a legitimate option in clinical practice) is the second key attribute of CER. This attribute also addresses the goal of putting new technology into proper perspective in relation to older technology. The importance of comparing alternative healthcare options is highlighted by the 2007 Institute of Medicine (IOM) report [44], which points out that "the rate with which new interventions are introduced into the medical marketplace is

Table 26.2 Attributes and corresponding goals of comparative effectiveness research studies.

Desired/necessary attributes	Corresponding CER goals
Real-world patients and settings	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients.
Head-to-head comparison of various treatment/diagnostic/ implementation strategies	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients. To put new technology into proper perspective versus older technology.
Inclusion of all stakeholders of healthcare (including patients/ caregivers) in the research process	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients.
Use clinically relevant and patient-centered outcomes	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients.
Assess heterogeneity of effects by patient variability, including phenotype and genotypes	To identify patients who are more or less likely to respond to some interventions than others.

currently outpacing the rate at which information is generated on their effectiveness and circumstances of best use" [3], and that "less than half of all medical care is based on or supported by adequate evidence about its effectiveness" [44]. In addition, wide variation in practice [45–49] as well as geographic variations in the utilization of certain treatments and procedures [50–55] suggest a lack of "sufficient evidence to determine which approach is most appropriate" [44].

The third key attribute invoked by the US CER initiatives is involving stakeholders, including patients and caregivers, in the research process [25,30,56]. As conceived in the IOM's recommendations for "a robust national CER enterprise" [25], this should involve a continuous process that considers and prioritizes topics for CER research and funding to address current knowledge gaps about diseases and conditions, and that consistently includes participation of patients, caregivers, and consumers to provide input regarding issues of public concern [25]. According to Slutsky *et al.* [30], priorities for CER must be based on input from all healthcare stakeholders, research and synthesis must apply to a wide range of healthcare services, and the results must be made accessible to multiple audiences. Stakeholders in healthcare are generally categorized as consumers (patients, caregivers, and the public), providers (clinicians), payers (health insurance programs and patients/ caregivers), policymakers, product makers (pharmaceutical industry), and researchers. A systematic review [57] assessing stakeholder engagement in CER and patient-centered outcomes research (PCOR) in published articles from 2003–2012 found that reports on stakeholder engagement were highly variable in content and quality. In this review, the most frequent engagement was with patients, engagement with clinicians was modestly frequent, and engagement with other groups was infrequent. Stakeholder engagement was more

common in the prioritization of CER than in its implementation and dissemination.

The fourth attribute of CER is that effectiveness and safety should be addressed using outcomes of interest and importance to patients and clinicians. This attribute also addresses the gap of traditional evidence based primarily on efficacy trials, many of which focused on surrogate outcomes instead of hard clinical outcomes or patient-reported outcomes [58] of most interest to patients and clinicians, such as quality of life or functional status.

Traditional efficacy trials typically report average effects and are usually underpowered to detect variability in patient responses. However, clinicians must make decisions about choices for patients whose profiles are similar to the average of study participants in the trials. Therefore, the fifth attribute of CER is exploration of heterogeneity to identify subgroups of patients who benefit more (or less) from a given intervention. While CER explores patient variability, it assesses treatment effects in subgroups that are not typically narrow enough to reflect differences in how individual patients respond to therapies [59]. Better practices are needed to evaluate treatment heterogeneity, accounting for more precise individual-level factors and preferences as well as genetic information, such as the conditional average treatment effect [59]. Developments in molecular biology and genomics will increasingly make it possible to assess genetic variation in individual responses to different treatment interventions, with the goal of individualized and predictive medicine [25] (see Chapter 30).

Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

All three components of CER (evidence synthesis, evidence generation, and evidence dissemination) are relevant in the field of pharmacoepidemiology. This section will cover all three components but will focus most on evidence generation, as it is the core field in pharmacoepidemiology and its methods are directly relevant to CER.

Issues for Evidence Synthesis

Systematic Reviews and Meta-Analyses in CER

The synthesis of evidence features prominently in definitions of CER, and systematic reviews and meta-analyses are the central approaches in evidence synthesis. To clarify common terminology following the Cochrane Collaboration and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [60,61], a systematic review refers a collection of all empirical evidence that fits prespecified eligibility criteria to answer a specific research question. Meta-analysis is the use of statistical methods to summarize and combine the results of independent studies (see Chapter 36). Therefore, many systematic reviews contain meta-analyses, but not all. Systematic review and meta-analyses can be used to discover patterns among study results and to provide reproducible summaries of study findings. In CER, systematic reviews may provide direct answers to CER questions, or may elucidate the need for more evidence generation when the results from individual studies are contradictory or when the magnitude of the underlying risk is small. Also, systematic reviews have been used in combination with clinical guidelines as a framework to identify knowledge gaps and to set research priorities [62].

The strengths of meta-analyses are mitigated by several methodologic challenges. The methodologic issues of systematic reviews and meta-analyses described in Chapter 36 are also relevant in evidence synthesis in CER. Briefly, the results of a meta-analysis are often highly subject to decisions made by the investigator: which studies to include or exclude from a meta-analysis, which outcome endpoints to

consider, and how to pool studies that differ in design and methods. Simmonds *et al.* [63] identified several sources of disagreement among experts that can affect the summary findings of meta-analyses. In addition, any limitations of the original studies will influence conclusions from the analysis of the pooled studies. This can be problematic in meta-analyses of observational studies. Consequently, some researchers have argued that only randomized trials should be meta-analyzed [64–67] (further discussion about meta-analyzing observational studies is found later in this chapter). It has also been argued that the outputs of meta-analyses may not provide greater insights than the results of individual studies [63,68].

Another limitation is that reviewers of the same studies may reach different conclusions, because of varying expertise in the topic of the review or in the technical skill of performing meta-analyses [69], or because of differences in values and orientations held by different investigators. The value of meta-analyses may also be seriously limited by publication bias, which can take several forms [70]. Studies with statistically nonsignificant or negative results are less likely to be published, and studies with statistically significant results and with stronger treatment effects tend to be published with less delay than studies with nonsignificant results. In addition, findings in some areas of research, such as complementary and alternative medicine, are less likely to be published. The summary conclusions from pooled published results will thus tend to be biased because of this preferential selectivity [70]. Problems stemming from publication bias may be amplified in meta-analyses of observational data. In addition, meta-analyses commonly combine the summary statistics from individual studies, whereas stronger results could be produced by obtaining and aggregating individual patient data from the separate studies analyzed [63,71,72]. However, issues of access, privacy, and ownership of original data make it difficult for investigators to obtain individual-level data.

Meta-Analyses of Observational CER

Generated evidence for CER can take the form of either observational studies or clinical trials. Therefore, longstanding debates about meta-analysis of observational studies are particularly relevant to CER. While some commentators have argued that meta-analyses of randomized trials are preferred to meta-analyses of observational studies [64–67], meta-analyses of observational studies are as common as those of randomized trials [68,73]. Reviews and practice guidelines on meta-analyzing observational data [68,73,74] show some disagreement with regard to this message. Some common opinions distilled include:

- Observational studies are more diverse in their designs and populations.
- Publication bias may be more problematic in observational studies [75,76].
- Biases are more problematic in observational studies.
- Therefore potential biases in the original studies make the calculation of a single summary estimate of effect of exposure potentially misleading, creating more precise but equally spurious effect estimates.
- More is gained by carefully examining possible sources of heterogeneity between the results of different observational studies.
- Concerns related to methodology and interpretation make the clear and thorough reporting of meta-analyses of observational studies absolutely essential (one guide provides a draft checklist summarizing recommendations for reporting meta-analyses of observational studies [73]).

From the point of view of pharmacoepidemiology, *a priori* exclusion of observational studies from meta-analyses would constitute a major loss. In fact, in some circumstances, meta-analysis of observational studies may be the only option to quantitatively synthesize current evidence. For example, Man *et al.* investigated the long-term effectiveness of methylphenidate in

the reduction of physical injuries. Harm reduction is a very important clinical outcome for patients and the healthcare system due to the high personal and economic cost of injuries. No clinical trials for methylphenidate were sufficiently long or even measured this outcome; hence, meta-analysis of observational studies was the only available option [77]. A study by Kirtane *et al.* [78] provides an example of a comprehensive meta-analysis that included both RCTs and observational studies, but analyzed them separately because of the differences in these types of study designs. Regardless of which types of studies are included in a meta-analysis, we agree with the need for a careful and systematic examination and reporting of observational studies, and for using epidemiologists' and clinicians' judgments to reach decisions about whether meta-analyses should be performed, and, if so, what studies should be included.

The expertise and effort required to perform a well-conceived and credible meta-analysis are not trivial. The AHRQ and IOM have published recommended standards for performing and reporting systematic reviews [79,80]. Nevertheless, the conclusions obtained by a rigorous meta-analysis cannot be deemed to provide a lasting answer to a clinical question, because new information may continuously become available. Therefore, the meta-analysis will require regular updates to keep it relevant for clinical guidelines [81].

Issues for Evidence Generation

Observational (Nonexperimental) Studies in CER

Observational studies have an important place in CER. First, observational studies provide data on real-world patients in usual clinical practice, which is one of the required attributes of CER evidence. Second, observational studies can provide larger samples and/or longer follow-up more easily than experimental studies. These are features that will be needed, as CER compares two options head to head, and this type of comparison will result in smaller effect sizes than comparing one treatment to placebo [42,82–84]. To date, the majority of CER studies have been conducted using observational study designs. Observational study designs and the methodologic issues they raise [85] are directly applicable to CER. In this section, we will summarize methodologic issues of particular importance in observational CER.

Confounding by Indication

As mentioned earlier, observational CER studies of intended effects are more susceptible to confounding by indication than observational studies of unintended effects (e.g., studies evaluating adverse drug events). While confounding by indication is covered in greater detail in Chapter 43, this bias is especially prominent when studying beneficial effects of treatments. In clinical practice, if one assumes prescribers are rational, one would expect treated patients to differ from untreated patients, as the former have an indication for the treatment. To the extent that the indication is related to the outcome variable as well, the indication can function as a confounding variable. On the other hand, confounding by indication for the treatment is less of a problem when a study is focused on unintended drug effects (side effects), regardless of whether those effects are harmful or beneficial. In this situation, the indication for treatment is less likely to be related to the outcome variable under study. However, this is sometimes not the case for studies of intended beneficial effects, which are the focus of many CER studies.

Confounding by indication (for the treatment) may also appear to be less of a problem when making comparisons between therapeutic alternatives for the same condition, since both study groups have the indication for treatment. However, nonrandomized studies comparing therapeutic alternatives are not necessarily free from confounding by indication, because the

true indication for a given treatment is often more subtle than the regulator-approved indication. For example, patients prescribed an angiotensin-converting enzyme (ACE) inhibitor as initial treatment for hypertension are likely to be different from those prescribed a thiazide diuretic for the same condition, as the former are more likely to have diabetes with nephropathy, myocardial infarction, and heart failure, whereas the latter being are more likely to have uncomplicated hypertension. As a second example, patients prescribed a combination of methotrexate and tumor necrosis factor alpha inhibitor as initial treatment for rheumatoid arthritis are likely to have more severe and active disease than those prescribed methotrexate monotherapy. Unless the choice between treatment alternatives is effectively random given measured variables, confounding by indication remains an issue in comparative studies. Disease severity is often associated with the outcomes of interests in studies assessing intended effects of treatment (e.g., improvements in rheumatoid arthritis symptoms or disease activity in the second example). Therefore, inability or limited ability to control for disease severity will results in bias (confounding by severity). The subtler examples of confounding by indication in the aforementioned scenarios are directly pertinent in CER, as these are exactly the types of questions CER addresses (comparing alternative options head to head). Furthermore, CER studies generally aim to detect differences that are likely to be smaller than in studies comparing exposed subjects to unexposed subjects. Accordingly, subtle instances of confounding by indication or confounding by severity can be especially problematic in CER.

Considerable effort has been undertaken to develop more effective methods for control of confounding in studies based primarily on administrative data (see also Chapters 12 and 43). Common approaches include propensity score-based methods, disease risk score-based methods, doubly robust methods, and instrumental variable analyses. These methods may, under certain conditions, provide better control of confounding than standard multivariable adjustment. However, it is important to keep in mind that most of these approaches (including propensity score but not instrumental variable analysis) are dependent on identifying and measuring those variables that are the true predictors of therapeutic choice in the databases.

For example, we conducted a study assessing the comparative effectiveness of carotid stenting versus carotid endarterectomy in older patients with carotid stenosis. Substantial confounding by indication would be suspected in this study, as carotid stenting is indicated and reimbursed only for patients with high surgical risks due to age, anatomic characteristics of carotid stenosis, or other cardiovascular or noncardiovascular co-morbidities (which are the predictors of worse prognosis). In this study, we demonstrated that propensity score-based methods (including high-dimensional propensity score methods [86]) using only claims data are insufficient to control for confounding by indication, and additional clinical information from vascular registries was necessary to achieve adequate control of confounding [87]. While these approaches provide sufficient confounding control in certain situations, they are generally not sufficient unless important variables related to the indication for treatment are available in the data [88]. Instrumental variables are promising alternatives if a valid instrument can be found for the clinical question. However, finding valid instruments in pharmacoepidemiology is extremely difficult – some would say impossible [89]. Design-based approaches such as restriction [90,91] and use of active comparators [92] may work in certain situations, and self-controlled methods can sometimes help to control time-invariant confounders [93], but it must be kept in mind that such approaches are directly linked to how the research question is defined. Accordingly, investigators must ensure

that the research question and the design are consistent and as intended. Much more work is needed in these areas to advance the field of CER.

Healthy User/Candidate Bias

Healthy user effect or bias has been observed among users of some medications, especially preventive medications such as hormone replacement therapies, statins, and certain antihypertensive medications [94–99]. For example, cardiovascular event reduction is consistently smaller in clinical trials compared with observational studies of antihypertensive medications [100,101], which suggests healthy user bias plays a role in observational studies of these preventive therapies. In CER comparing medical devices or interventions to a pharmacologic treatment, the healthy user effect (or more precisely for interventions, the healthy candidate effect) will be a major concern, as interventions typically pose short-term risks in exchange for long-term benefits, and patients at high risk of complications or deemed too sick to benefit are thus less likely to be selected for interventions. The healthy candidate effect is one of the biggest threats to validity in CER when comparing different treatment modalities. For example, we previously demonstrated the existence of healthy candidate bias in older heart failure patients who received implantable cardioverter-defibrillators (ICDs) versus those on medical therapies only by showing that patients with implanted ICDs had drastically better outcomes that were very unlikely to be attributed to the effects of ICDs (e.g., nursing home admissions, hip fracture, and short-term mortality) [102]. This highlights the utility of including falsification outcomes in observational studies.

Healthy user or candidate bias is thought of as a mix of confounding and selection bias. It arises when users of certain medications or candidates for invasive interventions have better outcomes due to factors other than effects of the treatment. While the factors associated with healthy user/candidate bias are not fully understood, many factors suggested to be responsible for healthy user/candidate effects are typically unmeasured in most databases. These include healthier lifestyles (healthy diets, regular exercise, and being less prone to using tobacco or alcohol) [99,103], higher socioeconomic status, better adherence to screenings and other preventive therapies [96], better physical [104] and cognitive function, less frailty [105], better social support, and stronger willingness to live. The effect size of healthy user/candidate bias can be quite substantial and is often as strong or stronger than the effect of the treatment itself [94–97,106]. Most importantly, healthy user bias may be refractory to analytic solutions unless prevented by thoughtful study design (e.g., self-controlled) and/or availability of extensive data on lifestyle and behavioral factors, thus resulting in inflation of the apparent benefits of preventive and other medications or invasive interventions.

Data Sources, Record Linkage, and Multidatabase Studies

As noted by the IOM report [25], CER studies should rely on multiple types of data sources, including primary data sources (medical and pharmacy records, electronic medical records, and *de novo* data generated through clinical trials or observational studies) and secondary data sources (administrative claims and clinical registries). Most CER studies to date have used the same data resources described in Chapters 12–14 of this book. As in usual pharmacoepidemiology practice, the data sources should be selected based on the study question and to maximize the internal and external validity of the results. To overcome the biases mentioned as well as others, including misclassification bias and selection bias, linking multiple data sources through record linkage can be a powerful tool, as it enriches the information for the given study subjects. Also, multidatabase studies within or across countries can potentially

enhance observational CER studies by enlarging the sample size for statistical power, assessment of effect heterogeneity, and improving generalizability. The methods, applications, and challenges of record linkage and multidatabase studies are described in other chapters.

A distinction must be made between data or record linkage (linking multiple data sources to enrich information) and multidatabase studies (using multiple databases for mostly nonoverlapping individuals). These two approaches are often confused, but each has a distinct goal. Data linkage is conducted in order to enrich information using a record linkage method, a computer-based technique to identify and link records from different databases that refer to the same entity or individual [107]. The data required for impactful and valid CER studies may be spread across two or more databases. Linking records across databases can transform ordinary individual datasets into powerful new platforms from which to perform timely and valid CER. For example, linkages between administrative claims databases and clinical or device registries can add longitudinal follow-up to registry data and add clinical details to administrative data. In addition to answering clinical questions, data linkage can also be used to address various methodologic issues in CER. For example, a linked database can be used to study data quality (e.g., by assessing agreement between two sources of the same data) and to validate claims-based endpoint ascertainment algorithms (e.g., by comparing a claims-based variable to a clinical gold standard) [108]. In addition to facilitating observational CER, record linkage can improve randomized trial evidence by linking patients in the trial to complementary data. For example, linking patients in a trial to Medicare claims can be a relatively inexpensive and effective way to extend the follow-up period of a clinical research study. Data linkage combining two or more data sources has enabled the conduct of more observational CER and/or more valid observational CER that would not be possible using a single data source [106,109–111]. Recent development of data linkage in Scandinavian countries has provided exciting opportunities to evaluate effectiveness beyond medical care. For example, Lichtenstein *et al*. linked the use of attention deficit hyperactivity disorder (ADHD) medication with criminal justice system records. They found that among patients with ADHD, rates of criminality were lower during periods when they were receiving ADHD medication. These findings raise the possibility that the use of medication reduces the risk of criminality among patients with ADHD [112].

The challenges of data linkage are both methodologic and ethical. Methodologic challenges include unavailability of linkage variables to researchers, especially unique identifiers such as names or social security numbers, incompleteness or inaccuracy of linkage variables due to poor data quality, nonoverlap or relatively small overlap of populations covered in each database, general misconceptions about linkage methods (especially probabilistic linkage methods), and understanding when to use what linkage method [113,114]. When unique identifiers of subjects are not available, at least for certain databases and populations (e.g., linking inpatient or outpatient claims data to clinical registry data for patients with heart failure, device implantations or surgeries, rheumatoid arthritis, and atrial fibrillation) [115,116], it is possible to conduct record linkage with high accuracy using multiple nonunique identifiers [115,116]. The primary ethical challenges of data linkage are ensuring patient privacy, which can be achieved by removing or limiting access to patient identifiers for research use including record linkage. However, this can make the linkage more difficult and sometimes impossible.

In our recent study of patients in a US-based online community, most reported that they were comfortable with researchers accessing their de-identified data for research purposes. Our study indicated that patient comfort levels

may be improved by better communication and transparency around specific research goals and how they may be beneficial to patient communities. In addition to mitigating re-identification risk, developing and improving methods to link databases through use of multiple nonunique identifiers may also improve patient comfort with secondary use of health data for research [117]. In a survey commissioned by the Wellcome Trust (a UK medical charity), 53% of respondents in the UK indicated that they would be happy for their data to be used by commercial organizations if it was for research purposes. Interestingly, over 60% of respondents indicated they would prefer that commercial research organizations have access to health data than that society miss out on the benefits these companies could potentially create. One of the most significant findings from the survey is that respondents considered academic researchers, charities, and organizations working in partnership with the public sector to be the most acceptable users of health data [118]. Patients' understanding and perceptions of the use of health data are still evolving, and it is important to continue to maintain communication and transparency regarding the use of health data for research.

In the last 10–15 years, networks of national and multinational database studies relevant to CER have been established in the US, Canada, Europe, and Asia [119–124]. Multidatabase study networks have been used to conduct observational CER or to provide a platform for CER trials [125]. The advantages of multidatabase studies in CER are that they capture diverse patient populations and/or increase the number of patients to detect relatively small effect sizes that can be expected in head-to-head comparisons and to assess the heterogeneity of effects. In multidatabase studies, data linkage methods are not necessary, as their intention is usually to bring databases together for nonoverlapping populations. However, structure, governance, and methods to manage and conduct a study using data from multiple sites and to synthesize results are needed. The structure, governance, and methods have to meet data management policies and data safety and privacy standards that may be unique to each database and can vary substantially, especially in international contexts [126,127].

Many networks employ a distributed network approach with a common data model (CDM), where the ownership and management of the database are left with individual data partners participating in the network. This approach is often preferred, as it mitigates most of the ethical and political issues with data privacy, governance, and ownership. In this model, each database is converted using a CDM so that its structure and coding are fully standardized. Multiple CDMs have been developed and modified to date [122,128–130]. To conduct analyses, researchers create a single statistical program that can be run against any database in the network with minimum or no modification. Another common approach is a distributed network with a common protocol rather than a CDM. The Canadian Network for Observational Drug Effect Studies (CNODES) operates using this approach, which eliminates the need to convert data from each site to a CDM [123]. In a distributed network with a CDM, a standardized coding language and format are needed to permit identical computerized queries to be submitted and executed across data resources, as well as standardized formats for returning responses from different databases [128,131,132].

In any approach for conducting multidatabase studies, understanding and dealing with variability in results across databases are especially challenging, particularly when the data come from different countries or diverse geographic regions/populations with differing healthcare systems, policies, and patient and clinician behaviors [126]. When large or unexpected variability in the results from each database is observed, researchers must first exclude the

possibility that the observed variability is due to technical issues arising from mapping codes or converting to a CDM, and/or from biases that are unique to each database (e.g., poor data quality, existence of and/or lack of understanding of unique features or idiosyncrasies in the data). When this possibility has been excluded, considerations must still be given as to when it is appropriate to combine results from different databases. This is especially important in CER, as understanding heterogeneity of effects is a major attribute of CER, and combining results that exhibit significant variability is not desired. The methodologic issues involved in combining results for meta-analyses discussed earlier are directly applicable to multidatabase studies as well, since it is generally not possible to analyze patient-level data to synthesize the results from each database, due to concerns about data security and privacy and/or restrictions of policies for data access and use.

Common challenges for data linkage and multidatabase studies include (i) logistical problems in accessing data sources, including issues of ownership of data, infrastructure, governance, data security, and data privacy (see also Chapters 12–14); and (ii) the required familiarity with the logical organization and content of disparate databases, including features or quirks in the data that are unique to each database. Needless to say, the aforementioned methodologic issues in observational CER (e.g., confounding by indication) can also affect the conduct and validity of results in linked database or multidatabase studies.

Experimental Studies

As already discussed, the goals of CER are to fill gaps in evidence that is traditionally and heavily based on premarketing RCTs. Observational studies leveraging existing data sources or primary data collection can be used as a valid and more cost-efficient approach for CER when available data include the necessary fields, and/or when researchers employ design-based or analytic methods to overcome potential biases. However, there are situations where bias is intractable and randomized trial designs are needed to obtain valid results. Large simple trials such as pragmatic trials or cluster randomized trials can determine the effects of an intervention under the usual conditions in which it will be applied, and therefore can assess real-world treatment effectiveness [133]. For clinical trials to be used in CER, researchers must focus on using trial designs that are flexible, adaptive, pragmatic, practical, and efficient, in contrast to traditional randomized, blinded, placebo-controlled clinical trials [134–137] (see also Chapter 32 for a discussion of large simple trials).

Briefly, pragmatic clinical trials are intended to overcome the limitations of traditional RCTs in order to answer CER questions. Pragmatic trials include real-world patients such as those with co-morbid conditions and those from diverse demographic backgrounds [138], and providers from community settings instead of only tertiary settings. In pragmatic CER trials, comparator treatments should be those in use in clinical practice (rather than placebo controls), outcomes should be those that matter to patients and clinicians rather than investigators or drug companies, and variations in patient responses to the treatment (treatment heterogeneity) should be explored [134,137,139].

One example of a pragmatic trial for CER is the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) [140], a $120 million NIH-funded trial comparing three antihypertensive medications and evaluating more than 42 000 patients in 600 clinics and centers in the US, Canada, Puerto Rico, and the US Virgin Islands [141]. A more recent example of a pragmatic trial for CER is the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) study, which compared ziprasidone and olanzapine for their risk of nonsuicide mortality [142,143]. This large randomized pragmatic trial included

approximately 18 000 patients in 18 countries and cost $85 million.

As in observational CER studies, using clinical trial designs for CER presents its own challenges. First, demonstrating clinically meaningful effect sizes is often challenging for several reasons: (i) the liberal inclusion criteria needed to assure generalizability of the study groups, although this increased heterogeneity can decrease the probability of detecting a given treatment effect as statistically significant, requiring an even larger sample size [134]; (ii) head-to-head comparison of commonly used clinical strategies; and (iii) modern medical interventions showing less dramatic benefits. The effect sizes seen even in placebo-controlled trials have been decreasing over time, and this pattern is considered to stem from the increasing rarity of discoveries of transformational medical interventions [144]. To detect a relatively small effect size and evaluate long-term and clinically relevant outcomes including hard endpoints, larger samples make CER trials more expensive [145]. The emphasis on usual care settings and the less-controlled nature of the trials lead to problems that are well known in observational studies. For example, loss to follow-up and/or nonadherence over time can introduce bias [135]. The lack of blinding in pragmatic trials creates the potential for biased observations and a threat to internal validity [134]. Another similar limitation of this type of trial results from the flexible treatment protocols that are preferred, as they are closer to what happens in real-world settings. Specifically, pragmatic trials involve the participation of community providers in their usual practice. Accordingly, providers can vary the treatment process, dose, and regimen given to different patients, depending on differing responses to therapy over time. This flexibility permits assessment of the outcomes of the composite treatment, but not of particular components within the treatment process [135].

As shown in the earlier examples of CER, the high cost of pragmatic trials is a major obstacle. However, conducting a simple randomized trial in usual clinical care conditions using routinely collected data, what are called "electronic point-of-care trials," could minimize the cost burden [146]. Attempts to conduct such trials with information technology (IT) tools to facilitate timely and efficient point of care (POC) recruitment have been reported by UK researchers using the General Practice Research Database® (GPRD®) [146,147]. In the US, researchers at Veterans Affairs (VA) hospitals have a head start taking advantage of the VA's sophisticated electronic health records (EHR) system [148]. The UK experience showed that the recruitment of clinicians and patients was a major challenge: the investigators observed that the number of interested clinicians/practices dropped substantially with each stage of the governance process, including site contracts, local approval forms, web-based good clinical practice, and protocol training [146]. A successful implementation of electronic POC trials will require three conditions at minimum: (i) a well-connected community with a network of practices and patients; (ii) data and IT infrastructure that enables patient recruitments at POC and captures clinically important outcomes; and (iii) readiness of clinicians and patients to accept "randomization" in routine clinical practice when there was equipoise among therapeutic options. A recent review article described attempts to conduct POC trials and integrate comparative effectiveness trials into patient care, illustrating challenges and limitations specific to POC trials [149]. Obviously, use of EHR poses limitations on the questions that can be addressed, processes that can be implemented, and outcomes that can be assessed.

Choosing the Right Methods: Experimental vs. Observational CER

There are inherent limitations in both experimental and observational designs, as discussed in the previous sections. While an experimental design is generally accepted and considered to

yield higher internal validity than observational CER, one must carefully examine how either of these types of studies are conducted in order to assess validity (both internal and external), interpret results, and draw meaningful conclusions. Several considerations can guide the decision of which study design to use for CER. First, some interventions cannot be investigated with clinical trials because of ethical considerations, even though such trials may otherwise be preferred scientifically. Second, it is not practical to employ pragmatic trials for most CER research questions due to their prohibitively high cost. Thus, observational studies using the techniques of nonexperimental pharmacoepidemiology will continue to play a role in CER, because some questions cannot be answered in clinical trials or because observational studies provide a cost-effective approach (when they can provide results with high internal validity).

Finally, there are some research questions that cannot be answered in observational CER due to intractable biases that severely compromise the internal validity of the results. Questions related to bias are not binary (existence or nonexistence of bias) but rather quantitative (degree of bias), and researchers must consider all potential biases and their quantitative impacts on the results before conducting observational CER studies. As one of the goals of CER is to produce results that are applicable and generalizable to real-world patients and practices, attempts to achieve higher generalizability may compromise features that are favorable to internal validity [84]. Nonetheless, results from CER studies that have significant bias and therefore have poor internal validity cannot be generalized. Once observational CER studies are completed, when intractable and significant biases are suspected in the results, researchers must provide a fair and honest assessment of study validity and must attempt to publish the results including this assessment in order for the research community to learn from their experiences.

Published Guides for CER Studies for Evidence Generation

Standards for performing and reporting observational studies have been provided by several professional associations [150–152]. While these guides are not specific to CER, they are relevant and directly applicable to observational CER. In addition, several other guides specifically targeting observational CER have been published through initiatives of AHRQ, PCORI, other governments, and professional societies (e.g., International Society of Pharmacoepidemiology [ISPE], International Society of Pharmacoeconomics and Outcomes Research [ISPOR], American Heart Association [AHA]) [151–160].

A recent systematic review of these CER-related guidance documents assessed shared expectations for quality CER [161]. The review identified nine documents with over 300 recommendations for designing and conducting CER. The most frequently shared recommendations included transparency and adaptation for relevant stakeholders in the interpretation and dissemination of results. Other frequently shared CER methods recommendations included developing an *a priori* study design and operational definitions that allow for replication (n=8 documents); focusing on areas with gaps in current clinical knowledge that are relevant to decision-makers (n=7); assessment and discussion of validity of measures, instruments, and data (n=7); and clinically meaningful and objectively measured outcomes, including benefits and harms (n=7). Additional commonly shared recommendations included assessment for and strategies to minimize bias (n=6 documents), confounding (n=6), and heterogeneity (n=4). Pragmatism in the design of experimental CER trials has been widely discussed [139,162–177], and there are proposed tools to assess pragmatism in clinical trials that researchers and clinicians can use when designing or evaluating pragmatic trials, especially for CER [178–181].

Issues for Evidence Dissemination

The ultimate goal of CER to improve clinical care will not be achieved without successful dissemination and adaptation of CER evidence. Evidence dissemination has several distinct goals. One goal involves identifying priority topics, comprehensively identifying available information on these topics, and developing objective interpretations of the information [182] (for example, as provided by Cochrane Collaboration reviews [183]). The output from this research then becomes the source information for dissemination to clinicians, patients, and policymakers. This goal will be achieved through expanding efforts on systematic reviews and studies using novel research designs (see earlier discussion), focusing on the priority research areas that were identified by the IOM as having knowledge gaps.

Another goal involves knowledge translation; namely, using research findings as the basis for drafting clinical guidelines. Achieving this goal will require qualified review panels that have scientific and clinical expertise in the content areas of the topics for which guidelines are developed, and who can develop clinical practice guidelines. Ideally, clinical guidelines should also be both comprehensive for general patient care and specific for particular patient circumstances – a very demanding specification. Furthermore, to remain relevant, guidelines need to be updated periodically to incorporate new information about existing interventions and new treatments. The IOM recently proposed standards for developing trustworthy clinical practice guidelines [184].

A third goal involves knowledge exchange and utilization, achieved by the actual distribution of information and the education of clinicians, patients, and policymakers about current knowledge and best practices. This goal may be attained by more intensive use of technology and/or social interventions. Examples of such tools include computerized physician order entry (CPOE) systems, supplemented by computerized clinical decision support systems (CDSSs) that incorporate electronic reminders to comply with guidelines (e.g., reminders to perform screening tests or to order other tests or treatments, reminders to avoid co-prescribing interacting drugs, etc.). Other strategies for achieving knowledge exchange and knowledge utilization will require educating clinicians and patients about what treatments work best [2,26,185]. These efforts should include monitoring to ensure that information is integrated into the normal workflow and decision processes of clinicians and patients.

A final goal involves monitoring and assessment of whether these efforts translate into actual good practice and, if not, to identify which means of dissemination have a greater chance to create an impact. However, recent history suggests that scientific evidence is often slow to change clinical practice. For example, despite harms associated with overdiagnosis of prostate cancer with prostate-specific antigen screening [186], the test is widely utilized in general practice [187]. Also, after the aforementioned multimillion-dollar ALLHAT pragmatic trial showed that thiazide diuretics are more effective than ACE inhibitors or calcium channel blockers for patients with hypertension, no significant changes in practice were observed [188]. Timbie *et al.* [189] reviewed CER studies conducted in the 2000s, including the ALLHAT trial, and identified five root causes underlying the failure of many CE studies to alter patient care:

- Financial incentives, such as fee-for-service payment, that may go against the adoption of new CER evidence.
- Ambiguity or concerns about the validity of CER study results.
- Common cognitive biases, including confirmation bias [189], pro-intervention bias [190], and pro-technology bias [191] in the interpretation of new CER information.

- Failure of research to address the needs of end users of CER evidence (clinicians, patients, policymakers).
- Limited use of decision support tools by patients and clinicians.

The authors offered several suggestions that align with the four dissemination goals already described. In addition, they suggested that in developing guidelines based on CER evidence, adapting the standards proposed by IOM (one of which was that guideline development groups be "multidisciplinary and balanced" [192]) may overcome several of the root causes mentioned, such as ambiguity of CER results and cognitive bias in interpreting the evidence. Finally, the authors proposed that aligning the incentives of clinicians and patients by changing payment and insurance models may facilitate the adoption of CER evidence in clinical practice. However, a recent systematic review found that pay-for-performance programs in healthcare were associated with improved processes, but not patient outcomes [193].

Dissemination of CER evidence is a legal mandate of PCORI [194,195]. The federal AHRQ works to disseminate findings from patient-centered outcomes research funded by PCORI, as well as government agencies and other sources. PCORI's release of new evidence from the funded studies begins with translating all research findings into understandable summaries with the help of the Patient-Centered Outcomes Research Translation Center. PCORI funds support not only engagement activities and infrastructure development, but also research to bring findings from completed studies into practice in real-world settings, and to compare approaches to communicating and disseminating patient-centered outcomes research findings, as well as research on shared decision-making [194]. PCORI recently announced a dissemination initiative (Eugene Washington PCORI Engagement Awards) through which it was planning to award $20.5 million in fiscal year 2018. Between fiscal years 2011 and 2017, AHRQ committed about $260 million for the dissemination and implementation of CER findings.

CER and Cost-Effectiveness Analyses

The primary goal of CER is to inform decisions that lead to better care, not necessarily cheaper care [196]. This could result in abandoning expensive technologies that are no better than less expensive options. However, it could also result in paying for a more expensive technology because the evidence shows it is superior [33,196,197]. The relevance of including cost-effectiveness analyses in CER investigations (see Chapter 34) is unquestionable. However, CER should not be used for cost-containment decisions [33,40,198], and the experts conducting CER studies should not be placed in the position of using their findings about treatment effectiveness to make recommendations about reimbursement. Nevertheless, well-performed CER inevitably should and will affect reimbursement decisions. In some cases, CER studies will find that the more expensive treatment is preferable. Yet, over time, CER should ultimately save money by preventing wasteful spending on treatments that are less effective, especially if dissemination is successful and CER evidence is adapted into clinical practice [196]. Ultimately, CER alone will not solve the US's healthcare spending problem ($3.3 trillion in 2016 – 17.9% of GDP – and $5.7 trillion projected in 2026 – 19.7% of GDP [199]).

Currently Available Solutions

In a review of recently published studies of the comparative effectiveness of existing (rather than new) medications, Hochman and McCormick [200] compared active therapies to each other (rather than to placebo comparators); compared medications to nonpharmacologic

interventions such as surgery or lifestyle interventions; compared different pharmacologic strategies for medication use; and compared different medication doses, durations, or formulations. They found that only one-third of studies evaluating medications qualified as comparative effectiveness research, and only a minority compared pharmacologic and nonpharmacologic therapies, emphasizing the need to expand the scope of CER. Another study reviewed clinical trials conducted in the US between 2007 and 2010 addressing priority CER topics identified by the IOM [201]. Among 1035 studies found on clinicaltrials.gov, 231 (22%) were comparative effectiveness (CE) studies. The most common interventions examined in CE studies were drugs (37%), behavioral interventions (29%), and procedures (16%).

These studies show what is observed in major medical journals or on clinicaltrials.gov during the period 2007–2010, but more recent data are not available. As described in the next section, the predominant source of funding for CER in the US thus far has been the federal government. Since PCORI was established in 2010, it has funded 596 CER projects (approximately $1.7 billion) [5]. These include CER studies, projects to examine CER methods, and projects to build infrastructure for CER and PCOR. The most frequent disease conditions for funded studies include mental/behavioral health (115 studies), cancer (84), neurologic disorders (74), cardiovascular diseases (69), and multiple chronic conditions (58). Most-studied populations of interest include racial/ethnic minorities (290), individuals of low socioeconomic status (194), women (145), older adults (134), and patients with multiple chronic conditions (110). Most of PCORI's research projects awarded through fiscal year 2017 are still underway; only 53 projects had been completed at the end of fiscal year 2017, but many are projected to be completed between 2018 and 2020.

The initial wave of funding came from the President's budget proposal for fiscal year 2011, with $286 million for patient-centered health research (the rebranded term for CER) through AHRQ. Much more substantial funding continued to come from the nongovernmental, nonprofit PCORI and its Patient-Centered Outcomes Research Trust Fund, established in 2010. As of January 2018, PCORI had brought a total investment of over $2 billion in projects meeting its congressional mandates, including funding for nearly 400 CER studies ($1.7 billion), as well as projects to improve the methods ($129 million) and infrastructure for CER, including the National Patient-Centered Clinical Research Network (PCORnet; $374 million). PCORI is projected to commit an additional $721 million for awards in fiscal years 2018 through 2021 [5]. From fiscal years 2011 through 2017, HHS including AHRQ committed approximately $448 million from the Trust Fund. Of this amount, HHS committed approximately $260 million (or 58%) to the dissemination and implementation of CER findings. HHS is projected to commit an additional $120 million for these activities in fiscal years 2018 through 2020 [5].

The EU does not have a central budget for healthcare expenditure. However, the European Commission has provided funding to conduct studies in European countries via its research budget (e.g., the CEPHOS-LINK Project [202]). In individual countries such as England, the NIHR has also provided funding for CER, particularly via the HTA Programme [7]. Examples from Asia include the Hong Kong Government's funding of CER via the Health and Medical Research Fund [203].

The Future

Funding

Thus far, the predominant source of funding for CER in the US has been the federal government, whereas much of the funding for clinical efficacy

research (i.e., RCTs) comes from industry sources. Recently, real-world evidence (RWE) has been gaining in popularity in the US, since the FDA's leadership published its opinion [204] and guidance document [205] on RWE, defining terminology and discussing its use in regulatory decision-making. RWE is defined by the FDA as "the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of Real-World Data (RWD)" [206]. In the EU, according to the GetReal Glossary of Definitions of Common Terms [207], "RWE derives from the analysis and/or synthesis of RWD that can either be primary data collected in a manner which reflects how interventions would be used in routine clinical practice or secondary data derived from routinely collected data." There is substantial overlap between RWE and "evidence generation in CER," as RWE includes comparative evidence of the "potential benefits and risks of medical products."

With the emphasis on RWE among regulatory bodies in the US and Europe, there has been increasing interest in RWE/CER among producers of medical products. In this new CER/RWE environment where the results of RWE or CER studies can be used in regulatory decision-making, companies may invest resources in RWE/CER, thus allowing them to compete on the utility (e.g., comparative effectiveness) of their products rather than just on their marketing ability [196]. For example, two global pharmaceutical companies recently collaborated to invest in a set of RWD analyses, including CER studies, of direct oral anticoagulants [208,209]. As most head-to-head comparative randomized studies are already sponsored by industry [210], with the increasing demand for RWE in North America, Europe, and Asia, future comparative studies may be conducted using RWD to provide CER evidence for new medical products. In addition to the possibility of more funding from industry for CER, public funding agencies and charities in the US and Europe have supported and will continue to support nonpharmacologic trials and studies of off-patent medications, such as through the Better Medicines for Children Initiative [211].

Human Capital Development

As noted by the Federal Coordinating Council [26], training will be required for new researchers to apply the specialized methods of CER and to develop new CER methods [212]. Specialized skills are needed to perform both traditional RCTs and novel pragmatic trials. It should be noted that these are not necessarily new but specialized skills, as CER is not a novel concept. Specialized expertise is also needed to perform formal meta-analyses and nonexperimental studies, using either *ad hoc* data collection or existing databases, and to successfully access and link various databases and conduct multidatabase studies. Finally, the field needs individuals who are able to translate the findings into practice guidelines and for other dissemination channels. The emphasis of CER on community participation and inclusion will dictate that experts from many different fields and backgrounds will be required to communicate with each other, finding and developing a common language to permit productive interactions. Therefore, the research teams participating in CER will be composed of professionals from different disciplines and different settings, including pharmacoepidemiologists and practicing clinicians from specialties relevant to the given studies [40]. These teams will need to have the capacity to develop a shared understanding of basic scientific terminology and methods.

Accordingly, it is necessary to create and support training programs for researchers seeking careers in CER in order to develop capacity in the research community to conduct CER. In addition to preparing a cadre of researchers with expertise in CER methods, a critical mass of such researchers is required in order to undertake the large number of studies needed

to fill current knowledge gaps and to continuously update the knowledge inventory with systematic reviews. So far, PCORI has committed $30 million to workforce training awards for clinicians and researchers. For example, one of its career development programs, conducted in partnership with AHRQ, is designed to train clinician and research scientists to conduct patient-centered outcomes research and to actively engage stakeholders in efforts to improve the quality and safety of care [5]. AHRQ committed $94 million for efforts to train researchers on the conduct of CER [5] and plans to commit an additional $14 million by fiscal year 2020 for CER training. In Europe, the European Commission has provided funding for training in health and medical research including CER and pharmacoepidemiology in European countries (e.g., the Marie Skłodowska-Curie Fellowship [213]). In England, the NIHR has also provided funding for fellowships in health research including CER research, particularly through the NIHR Fellowship Programme [214]. In Asia, the Hong Kong Government funds health fellowships via the Health and Medical Research Fund [203].

CER and Clinical Practice

Though the research community continues to appear excited and energized, expectations must be tempered by several limits on what CER can realistically solve. It is unrealistic to expect that CER will address all therapeutic questions; healthcare is simply too complex. Sir William Osler, the Father of modern medicine, wrote, "The practice of medicine is an art, based on science" [215]. Even in the era of evidence-based medicine and CER, the practice of medicine is as much the application of art as the application of evidence. To reach an optimal decision in any given clinical situation, evidence must be applied to an individual patient who has her or his own values, preferences, life situations, and goals. Treating not only the disease but also the patient as a whole requires both understanding and application of the best evidence, as well as the skills and behaviors physicians bring to their own practice of the art of medicine [216]. Alternatively, if it were possible to base medical practice entirely on evidence, such evidence would consider not only complex pathophysiology, but also personal factors such as values, preferences, perceptions, and attitudes about risks, quality-of-life preferences, cost tradeoffs, as well as clinician–patient interactions. However, as the evidence underlying current medical practice consists of estimated averages from studied populations and is also not sufficiently complete, art comes into practice when subjective judgment is required. Therefore, even in situations where complete evidence-based information is available to guide clinical decisions, providers or patients may still opt for a decision based on personal choices that they value irrespective of the scientific evidence. As stated by Kerridge *et al.* [217]:

> Medical decision making draws upon a broad spectrum of knowledge – including scientific evidence, personal experience, personal biases and values, economic and political considerations, and philosophical principles (such as concern for justice). It is not always clear how practitioners integrate these factors into a final decision, but it seems unlikely that medicine can ever be entirely free of value judgments.

Overemphasis on scientific evidence can lead to therapeutic nihilism; that is, paralysis when such evidence is unavailable. Also, an overreliance on evidence-based guidelines can result in algorithmic care [218]. Ironically, this in turn may devalue individualized care, which is another goal and feature of CER. Subgroup analysis is the most commonly used approach for assessing heterogeneity in CER, but it has obvious limitations in providing sufficient evidence to "individualize" the care of each

patient. In the face of uncertainty, variation among reasonable but unproven options should be tolerated and even encouraged, as it will facilitate later evaluation. This runs contrary to the vision of a knowledge state that is sufficiently complete to guide all decisions about effective interventions at the individual patient level. We also need to be sure that the desire for scientific evidence does not paralyze medical practice when such evidence is absent. In such circumstances, the resulting variability in practice can provide the data that will underlie future CER studies.

As healthcare communities continue to embrace CER and demand better evidence to inform clinical decisions, CER will play an expanding role in healthcare research. It continues to establish its position as a central component of clinical research that is directly relevant to clinical practice and health policy; that is, CER is needed in order to practice the best evidence-based medicine and evidence-based policymaking. Nonetheless, creating sustainable funding sources, improving the quality of evidence generation and evidence dissemination through development of methods and better use of these methods, educating consumers about CER, and managing their expectation are ongoing challenges.

References

1 Luce BR, Drummond M, Jönsson B, *et al*. EBM, HTA, and CER: clearing the confusion. *Milbank Q* 2010; **88**: 256–76.

2 Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. *JAMA* 2010; **303**: 985–6.

3 Congressional Budget Office. Research on the comparative effectiveness of medical treatments. Nov 2007. http://www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf (accessed January 2011).

4 H.R. 1 The Medicare Prescription Drug, Improvement, and Modernization Act of 2003. Public Law 108-173, enacted December 8, 2003.

5 United States Government Accountability Office. Comparative effectiveness research: activities funded by the Patient-Centered Outcomes Research Trust Fund. March 2018. https://www.gao.gov/assets/700/690835.pdf (accessed September 2018).

6 Patient-Centered Outcomes Research Institute. Research Funding. 2018. https://www.pcori.org/sites/default/files/PCORI-Research-Funding.pdf (accessed September 2018).

7 National Institute for Health Research. Health Technology Assessment. https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment (accessed October 2018).

8 Chalkidou K, Walley T. Using comparative effectiveness research to inform policy and practice in the UK NHS: past, present and future. *Pharmacoeconomics* 2010; **28**: 799–811.

9 National Institute for Health and Care Excellence. What we do. https://www.nice.org.uk/about/what-we-do (accessed October 2018)

10 National Institute for Health and Care Excellence. NICE guidance. https://www.nice.org.uk/guidance (accessed October 2018).

11 Poole-Wilson PA. NICE: guessing is not guidance. National Institute for Clinical Excellence. *Heart* 2002; **87**(6): 510–12.

12 National Institute for Health and Care Excellence. Ixazomib with lenalidomide and dexamethasone for treating relapsed or refractory multiple myeloma. Technology appraisal guidance [TA505]. February 2018.

13 National Institute for Health and Care Excellence. Neuropad for detecting preclinical diabetic peripheral neuropathy. Medical

technologies guidance [MTG38]. September 2018.
14 National Institute for Health and Care Excellence. Intravesical microwave hyperthermia and chemotherapy for non-muscle-invasive bladder cancer. Interventional procedures guidance [IPG628]. September 2018.
15 Sorenson C. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison. *Issue Brief (Commonw Fund)* 2010; **91**: 1–14.
16 Evans H. Comparative effectiveness in health care reform: lessons from abroad. Heritage Foundation. 2009. http://www.heritage.org/Research/Reports/2009/02/Comparative-Effectiveness-in-Health-Care-Reform-Lessons-from-Abroad (accessed October 2018).
17 Haute autorité de santé (HAS). https://www.has-sante.fr/portail/jcms/r_1455081/en/home-page?portal=r_1455081 (accessed May 2019).
18 Institute for Quality and Efficiency in Health Care. Responsibilities and objectives of IQWiG. https://www.iqwig.de/en/about-us/responsibilities-and-objectives-of-iqwig.2946.html (accessed October 2018).
19 Levy AR, Mitton C, Johnston KM, Harrigan B, Briggs AH. International comparison of comparative effectiveness research in five jurisdictions: insights for the US. *Pharmacoeconomics* 2010; **28**: 813–30.
20 Sweeny K. Key aspects of the Australian Pharmaceutical Benefits Scheme. Working Paper No. 35. Pharmaceutical Industry Project. November 2007. http://vuir.vu.edu.au/15944 (accessed May 2019).
21 Miller LJ, Lu W. These are the economies with the most (and least) efficient health care. https://www.bloomberg.com/news/articles/2018-09-19/u-s-near-bottom-of-health-index-hong-kong-and-singapore-at-top (accessed October 2018).
22 Chalkidou K. From methods to policy: comparative effectiveness research around the globe: a valuable tool for achieving and sustaining universal healthcare. *J Comp Eff Res* 2018; **7**(5): 503–15.
23 Patel I, Rarus R, Tan X, *et al*. Investigation of comparative effectiveness research in Asia, Europe, and North America. *Indian J Pharmacol* 2015; **47**(6): 585–93.
24 Congressional Budget Office. Research on the comparative effectiveness of medical treatments: issues and options for an expanded federal role. 2007. www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf (accessed October 2018).
25 Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington, DC: National Academies Press, 2009. http://www.nationalacademies.org/hmd/reports/2009/comparativeeffectivenessresearchpriorities.aspx (accessed December 2010).
26 Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. Washington, DC: Department of Health and Human Services, June 2009. https://osp.od.nih.gov/wp-content/uploads/FCCCER-Report-to-the-President-and-Congress-2009.pdf (accessed December 2010).
27 Patient-Centered Outcomes Research Institute. Research we support. https://www.pcori.org/research-results/research-we-support (accessed September 2018).
28 European Commission. Health technology assessment. https://ec.europa.eu/health/technology_assessment/overview_en (accessed October 2018).
29 International Network of Agencies for Health Technology Assessment. Welcome to INAHTA. http://www.inahta.org (accessed October 2018).
30 Slutsky JR, Clancy CM. Patient-centered comparative effectiveness research. *Arch Intern Med* 2010; **170**: 403–4.
31 Benner JS, Morrison MR, Karnes EK, *et al*. An evaluation of recent federal spending on

comparative effectiveness research: priorities, gaps, and next steps. *Health Aff (Millwood)* 2010; **29**: 1768–76.

32 Sox HC, Helfand M, Grimshaw J, *et al*. Comparative effectiveness research: challenges for medical journals. *Trials* 2010; **28**(11): 45.

33 Selker HP. Comparative effectiveness research: medical practice, payments, and politics: the need to retain standards of medical research. *J Gen Intern Med* 2009; **24**: 776–8.

34 Martin DF, Maguire MG, Fine SL. Identifying and eliminating the roadblocks to comparative-effectiveness research. *N Engl J Med* 2010; **8**(363): 105–7.

35 Slone D, Shapiro S, Miettinen OS, *et al*. Drug evaluation after marketing. *Ann Intern Med* 1979; **90**(2): 257–61.

36 Siskind V. The need for randomization in the study of intended effects. *Stat Med* 1985; **4**(3): 387–9.

37 Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med* 1985; **78**(3): 475–80.

38 Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**(4): 703–8.

39 Iglehart JK. Prioritizing comparative-effectiveness research: IOM recommendations. *N Engl J Med* 2009; **361**: 325–8. DOI:10.1056/NEJMp0904133

40 Manchikanti L, Falco FJ, Boswell MV, *et al*. Facts, fallacies, and politics of comparative effectiveness research: part I. *Basic considerations. Pain Physician* 2010; **13**: E23–E54.

41 Volpp KG, Das A. Comparative effectiveness: thinking beyond medication A versus medication B. *N Engl J Med* 2009; **361**: 331–3.

42 Lauer MS, Collins FS. Using science to improve the nation's health system: NIH's commitment to comparative effectiveness research. *JAMA* 2010; **303**: 2182–3.

43 Aston G. Comparative effectiveness (why does it matter to you?). *Hosp Health Netw* 2009; **83**: 22–5.

44 Institute of Medicine. Learning what works best: the nation's need for evidence on comparative effectiveness in health care. September 2007. https://www.ncbi.nlm.nih.gov/books/NBK64784 (accessed May 2019).

45 Institute of Medicine. Knowing what works in health care: a roadmap for the nation. January 2008. https://www.nap.edu/catalog/12038/knowing-what-works-in-health-care-a-roadmap-for-the (accessed May 2019).

46 Tang A, Hooper T, Hasan R. A regional survey of chest drains: evidence-based practice? *Postgrad Med J* 1999; **75**: 471–4.

47 Likosky DS, FitzGerald DC, Groom RC, *et al*. Effect of the perioperative blood transfusion and blood conservation in cardiac surgery clinical practice guidelines of the Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists upon clinical practices. *Anesth Analg* 2010; **111**: 316–23.

48 Goldberger JJ, Bonow RO, Cuffe M, *et al*. Beta-blocker use following myocardial infarction: low prevalence of evidence-based dosing. *Am Heart J* 2010; **160**: 435–42 e1.

49 Phillips KA, Marshall DA, Haas JS, *et al*. Clinical practice patterns and cost effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. *Cancer* 2009; **115**: 5166–74.

50 Piccini JP, Hernandez AF, Dai D, *et al*.; Get With the Guidelines Steering Committee and Hospitals. Use of cardiac resynchronization therapy in patients hospitalized with heart failure. *Circulation* 2008; **118**: 926–33.

51 Matlock DD, Peterson PN, Heidenreich PA, *et al*. Regional variation in the use of implantable cardioverter-defibrillators for primary prevention: results from the National Cardiovascular Data Registry. *Circ Cardiovasc Qual Outcomes* 2011; **4**(1): 114–21.

52 Sung JC, Curtis LH, Schulman KA, *et al*. Geographic variations in the use of medical and surgical therapies for benign prostatic hyperplasia. *J Urol* 2006; **175**(3 Pt 1): 1023–7.

53 Warren OJ, Wallace S, de Wit KL, *et al.* Variations in the application of various perfusion technologies in Great Britain and Ireland: a national survey. *Artif Organs* 2010; **34**: 200–5.

54 Lougheed MD, Garvey N, Chapman KR, *et al.* Variations and gaps in management of acute asthma in Ontario emergency departments. *Chest* 2009; **135**: 724–36.

55 Luo X, Pietrobon R, Curtis LH, *et al.* Prescription of nonsteroidal anti-inflammatory drugs and muscle relaxants for back pain in the United States. *Spine* 2004; **29**: E531–E537.

56 Wu AW, Snyder C, Clancy CM, *et al.* Adding the patient perspective to comparative effectiveness research. *Health Aff (Millwood)* 2010; **29**(10): 1863–71.

57 Concannon TW, Fuster M, Saunders T, *et al.* A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. *J Gen Intern Med* 2014; **29**(12): 1692–701.

58 Howie L, Hirsch B, Locklear T, *et al.* Assessing the value of patient-generated data to comparative effectiveness research. *Health Aff (Millwood)* 2014; **33**(7): 1220–8.

59 Basu A, Jena AB, Goldman DP, *et al.* Heterogeneity in action: the role of passive personalization in comparative effectiveness research. *Health Econ* 2014; **23**(3): 359–73.

60 Green S, Higgins JPT, Alderson P, *et al.* What is a systematic review? In *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.0.0. London: Cochrane Collaboration, 2008, Chapter 9.1. https://community.cochrane.org/book_pdf/764 (accessed May 2019).

61 Liberati A, Altman DG, Tetzlaff J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

62 Li T, Vedula SS, Scherer R, *et al.* What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. *Ann Intern Med* 2012; **156**(5): 367–77.

63 Simmonds MC, Higgins JP, Stewart LA, *et al.* Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2005; **2**(3): 209–17.

64 Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; **140**(3): 290–6.

65 Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; **351**(9096): 123–7.

66 Shapiro S. Meta-analysis/shmeta-analysis. *Am J Epidemiol* 1994; **140**(9): 771–8.

67 Blettner M, Sauerbrei W, Schlehofer B, *et al.* Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999; **28**(1): 1–9.

68 Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**(7125): 140–4.

69 Manchikanti L, Datta S, Smith HS, *et al.* Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 6. Systematic reviews and meta-analyses of observational studies. *Pain Physician* 2009; **12**: 819–50.

70 Song F, Parekh S, Hooper L, *et al.* Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010; **14**: iii, ix–xi, 1–193.

71 Broeze KA, Opmeer BC, Bachmann LM, *et al.* Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol* 2009; **9**: 22.

72 Koopman L, van der Heijden GJ, Hoes AW, *et al.* Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *Int J Technol Assess Health Care* 2008; **24**: 358–61.

73 Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**(15): 2008–12.

74 Shrier I, Boivin JF, Steele RJ, *et al*. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; **166**(10): 1203–9.

75 Easterbrook PJ, Berlin JA, Gopalan R, *et al*. Publication bias in clinical research. *Lancet* 1991; **337**(8746): 867–72.

76 Schlesselman JJ. Risk of endometrial cancer in relation to use of combined oral contraceptives: a practitioner's guide to meta-analysis. *Hum Reprod* 1997; **12**(9): 1851–63.

77 Man KKC, Ip P, Chan EW, *et al*. Effectiveness of pharmacological treatment for attention-deficit/hyperactivity disorder on physical injuries: a systematic review and meta-analysis of observational studies. *CNS Drugs* 2017; **31**(12): 1043–55.

78 Kirtane AJ, Gupta A, Lyengar S, *et al*. Safety and efficacy of drug-eluting and bare metal stents: comprehensive meta-analysis of randomized trials and observational studies. *Circulation* 2009; **119**: 3198–206.

79 AHRQ's methods research paper on comparative effectiveness review methods: clinical heterogeneity. AHRQ publication no. 10-EHC-070-EF. September 2018. https://www.ncbi.nlm.nih.gov/books/NBK53310 (accessed May 2019).

80 Institute of Medicine. Standards for systematic reviews of clinical effectiveness research. finding what works in health care: standards for systematic reviews. September 2018. https://www.ncbi.nlm.nih.gov/books/NBK209526 (accessed May 2019).

81 Shojania KG, Sampson M, Ansari MT, *et al*. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007; **147**: 224–33.

82 Dreyer NA, Tunis SR, Berger M, *et al*. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)* 2010; **29**: 1818–25.

83 Teutsch SM, Berger ML, Weinstein MC. Comparative effectiveness: asking the right questions, choosing the right method. *Health Affairs* 2005; **24**: 128–32.

84 Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health* 2012; **33**: 425–45.

85 Strom BL. Methodologic challenges to studying patient safety and comparative effectiveness. *Med Care* 2007; **45**(10 Suppl 2): S13–S15.

86 Schneeweiss S, Rassen JA, Glynn RJ, *et al*. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; **20**(4): 512–22.

87 Kumamaru H, Jalbert JJ, Setoguchi S. Incremental utility of additional clinical information and propensity score methods in controlling for severe confounding by indication. *Pharmacoepidemiol Drug Saf* 2018; **27**: 489.

88 Green SB, Byar DP. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med* 1984; **3**: 361–70.

89 Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–72.

90 Schneeweiss S, Patrick AR, Sturmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–S142.

91 McGrath LJ, Ellis AR, Brookhart MA. controlling time-dependent confounding by health status and frailty: restriction versus statistical adjustment. *Am J Epidemiol* 2015; **182**(1): 17–25.

92 Setoguchi S, Gerhard T. Comparator selection. In Velentgas P, Dreyer NA, Nourjah P, *et al.*, eds., *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: AHRQ, 2013, Chapter 5. https://www.ncbi.nlm.nih.gov/books/NBK126184 (accessed May 2019).

93 Lao KS, Chui CS, Man KK, *et al.* Medication safety research by observational study design. *Int J Clin Pharm* 2016; **38**(3): 676–84.

94 Majumdar SR, McAlister FA, Eurich DT, *et al.* Statins and outcomes in patients admitted to hospital with community acquired pneumonia: population based prospective cohort study. *BMJ* 2006; **333**(7576): 999.

95 Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002; **137**(4): 273–84.

96 Brookhart MA, Patrick AR, Dormuth C, *et al.* Adherence to lipid-lowering therapy and the use of preventive health services: an investigation of the healthy user effect. *Am J Epidemiol* 2007; **166**(3): 348–54.

97 Ray WA, Daugherty JR, Griffin MR. Lipid-lowering agents and the risk of hip fracture in a Medicaid population. *Inj Prev* 2002; **8**(4): 276–9.

98 Setoguchi S, Avorn J, Schneeweiss S. Statins and the risk of colorectal cancer. *N Engl J Med* 2005; **353**(9): 952–4.

99 Kinjo M, Chia-Cheng Lai E, Korhonen MJ, *et al.* Potential contribution of lifestyle and socioeconomic factors to healthy user bias in antihypertensives and lipid-lowering drugs. *Open Heart* 2017; **4**(1): e000417.

100 Blood Pressure Lowering Treatment Trialists Collaboration, Turnbull F, Neal B, *et al.* Effects of different regimens to lower blood pressure on major cardiovascular events in older and younger adults: meta-analysis of randomised trials. *BMJ* 2008; **336**(7653): 1121–3.

101 Mazzaglia G, Ambrosioni E, Alacqua M, *et al.* Adherence to antihypertensive medications and cardiovascular morbidity among newly diagnosed hypertensive patients. *Circulation* 2009; **120**(16): 1598–605.

102 Setoguchi S, Warner Stevenson L, Stewart GC, *et al.* Influence of healthy candidate bias in assessing clinical effectiveness for implantable cardioverter-defibrillators: cohort study of older patients with heart failure. *BMJ* 2014; **348**: g2866.

103 Halava H, Korhonen MJ, Huupponen R, *et al.* Lifestyle factors as predictors of nonadherence to statin therapy among patients with and without cardiovascular comorbidities. *CMAJ* 2014; **186**(12): E449–E456.

104 Dublin S, Jackson ML, Nelson JC, *et al.* Statin use and risk of community acquired pneumonia in older people: population based case-control study. *BMJ* 2009; **338**: b2137.

105 Glynn RJ, Knight EL, Levin R, *et al.* Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* 2001; **12**(6): 682–9.

106 Chen CY, Stevenson LW, Stewart GC, *et al.* Real world effectiveness of primary implantable cardioverter defibrillators implanted during hospital admissions for exacerbation of heart failure or other acute co-morbidities: cohort study of older patients with heart failure. *BMJ* 2015; **351**: h3529.

107 Christen P. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. New York: Springer, 2012.

108 Setoguchi S, Solomon DH, Glynn RJ, *et al.* Agreement of diagnosis and its date for hematologic malignancies and solid tumors between Medicare claims and cancer registry data. *Cancer Causes Control* 2007; **18**(5): 561–9.

109 Douglas PS, Brennan JM, Anstrom KJ, *et al*. Clinical effectiveness of coronary stents in elderly persons: results from 262,700 Medicare patients in the American College of Cardiology-National Cardiovascular Data Registry. *J Am Coll Cardiol* 2009; **53**(18): 1629–41.

110 Weintraub WS, Grau-Sepulveda MV, Weiss JM, *et al*. Comparative effectiveness of revascularization strategies. *N Engl J Med* 2012; **366**(16): 1467–76.

111 Jalbert JJ, Gerhard-Herman MD, Nguyen LL, *et al*. Relationship between physician and hospital procedure volume and mortality after carotid artery stenting among Medicare beneficiaries. *Circ Cardiovasc Qual Outcomes* 2015; **8**(6 Suppl 3): S81–S89.

112 Lichtenstein P, Halldner L, Zetterqvist J, *et al*. Medication for attention deficit-hyperactivity disorder and criminality. *N Engl J Med* 2012; **367**(21): 2006–14.

113 Zhu Y, Chen CY, Matsuyama Y, *et al*. Comparative validity of methods to select appropriate cutoff weight for probabilistic linkage without unique personal identifiers. *Pharmacoepidemiol Drug Saf* 2016; **25**(4): 444–52.

114 Zhu Y, Matsuyama Y, Ohashi Y, *et al*. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J Biomed Inform* 2015; **56**: 80–6.

115 Setoguchi S, Zhu Y, Jalbert JJ, *et al*. Validity of deterministic record linkage using multiple indirect personal identifiers: linking a large registry to claims data. *Circ Cardiovasc Qual Outcomes* 2014; **7**(3): 475–80.

116 Curtis JR, Chen L, Bharat A, *et al*. Linkage of a de-identified United States rheumatoid arthritis registry with administrative data to facilitate comparative effectiveness research. *Arthritis Care Res (Hoboken)* 2014; **66**(12): 1790–8.

117 O'Brien E, O'Brien S, Setoguchi S. Patient perspectives on the linkage of health data for clinical research: insights from a survey in the United States. *World Congress of Epidemiology*. International Epidemiological Association. Saitama, Japan, August 2017.

118 Ipsos MORI. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. March 2016. https://wellcome.ac.uk/what-we-do/our-work/our-policy-work-using-patient-data-research (accessed October 2018).

119 Coloma PM, Schuemie MJ, Trifiro G, *et al*. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011; **20**(1): 1–11.

120 Platt R, Wilson M, Chan KA, *et al*. The new Sentinel Network: improving the evidence of medical-product safety. *N Engl J Med* 2009; **361**(7): 645–7.

121 Andersen M, Bergman U, Choi NK, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; **22**(7): 700–4.

122 Lai EC, Ryan P, Zhang Y, *et al*. Applying a common data model to Asian databases for multinational pharmacoepidemiologic studies: opportunities and challenges. *Clin Epidemiol* 2018; **10**: 875–85.

123 Suissa S, Henry D, Caetano P, *et al*. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med* 2012; **6**(4): e134–e140.

124 PCORnet Front Door. https://pcornet.org/frontdoor (accessed May 2019).

125 ADAPTABLE. The Aspirin Study. PCORnet demonstration trial. http://theaspirinstudy.org (accessed May 2019).

126 Lai E, Stang P, Kao Yang Y, *et al*. International multi-database pharmacoepidemiology: potentials and

pitfalls. *Curr Epidemiol Rep* 2015; **2**(4): 229–38.

127 Lai EC, Shin J, Kubota K, *et al*. Comparative safety of NSAIDs for gastrointestinal events in Asia-Pacific populations: a multi-database, international cohort study. *Pharmacoepidemiol Drug Saf* 2018; **27**(11): 1223–30.

128 Sentinel Initiative. Distributed database and common data model. https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model (accessed September 2018).

129 Observational Health Data Sciences and Informatics. OMOP common data model. https://www.ohdsi.org/data-standardization/the-common-data-model (accessed September 2018).

130 PCORnet. PCORnet Common Data Model (CDM). https://pcornet.org/pcornet-common-data-model (accessed September 2018).

131 Maro JC, Platt R, Holmes JH, *et al*. Design of a national distributed health data network. *Ann Intern Med* 2009; **151**(5): 341–4.

132 Chan KA, HMO Research Network CERT Patient Safety Study Investigators. Development of a multipurpose dataset to evaluate potential medication errors in ambulatory settings. In Henriksen K, Marks ES, Lewin DI, eds, *Advances in Patient Safety: From Research to Implementation. Vol. 2: Concepts and Methodology*. AHRQ Publication No. 05-0021-2. Rockville, MD: AHRQ, 2005, pp. 225–38.

133 Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967; **20**(8): 637–48.

134 Macpherson H. Pragmatic clinical trials. *Complement Ther Med* 2004; **12**: 136–40.

135 Relton C, Torgerson D, O'Cathain A, *et al*. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *BMJ* 2010; **340**: c1066.

136 Brass EP. The gap between clinical trials and clinical practice: the use of pragmatic clinical trials to inform regulatory decision making. *Clin Pharmacol Ther* 2010; **87**: 351–5.

137 Alford L. On differences between explanatory and pragmatic clinical trials. *New Zeal J Physiotherapy* 2006; **35**(1): 12–16.

138 Tinetti ME, Studenski SA. Comparative effectiveness research and patients with multiple chronic conditions. *N Engl J Med* 2011; **364**(26): 2478–81.

139 Luce BR, Kramer JM, Goodman SN, *et al*. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med* 2009; **151**: 206–9.

140 Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002; **288**(23): 2981–97.

141 Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) Data Coordinating Center. Dose of reality: study questions high-cost drugs for hypertension. https://ccct.sph.uth.tmc.edu/ALLHAT/News/HighCostDrugs.aspx (accessed September 2018).

142 Strom BL, Eng SM, Faich G, *et al*. Comparative mortality associated with ziprasidone and olanzapine in real-world use among 18,154 patients with schizophrenia: the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC). *Am J Psychiatry* 2011; **168**: 193–201.

143 Strom BL, Faich GA, Reynolds RF, *et al*. The Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC): design and baseline subject characteristics. *J Clin Psychiatry* 2008; **69**: 114–21.

144 Olfson M, Marcus SC. Decline in placebo-controlled trial results suggests new directions for comparative effectiveness

research. *Health Aff (Millwood)* 2013; **32**(6): 1116–25.

145 Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *Br Med J (Clin Res Ed)* 1984; **289**: 1281–4.

146 van Staa TP, Dyson L, McCann G, *et al*. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014; **18**(43): 1–146.

147 van Staa TP, Klungel O, Smeeth L. Use of electronic healthcare records in large-scale simple randomized trials at the point of care for the documentation of value-based medicine. *J Intern Med* 2014; **275**(6): 562–9.

148 Fiore LD, Brophy M, Ferguson RE, *et al*. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials* 2011; **8**(2): 183–95.

149 Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. *N Engl J Med* 2016; **374**(22): 2152–8.

150 ISPE. Guidelines for Good Pharmacoepidemiology Practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.

151 Berger ML, Mamdani M, Atkins D, *et al*. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part I. *Value Health* 2009; **12**: 1044–52.

152 Cox E, Martin BC, Van Staa T, *et al*. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value Health* 2009; **12**: 1053–61.

153 Velentgas P, Dreyer NA, Nourjah P, *et al*., eds. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: AHRQ, 2013. https://www.ncbi.nlm.nih.gov/books/NBK126190 (accessed May 2019).

154 PCORI (Patient-Centered Outcomes Research Institute) Methodology Committee. The PCORI methodology report. https://www.pcori.org/research-results/about-our-research/research-methodology/pcori-methodology-standards (accessed September 2018).

155 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology (Revision 3). EMA/95098/2010. 2014. http://www.encepp.eu/standards_and_guidances (accessed October 2018).

156 Gibbons RJ, Gardner TJ, Anderson JL, *et al*. The American Heart Association's principles for comparative effectiveness research: a policy statement from the American Heart Association. *Circulation* 2009; **119**(22): 2955–62.

157 Johnson ML, Crown W, Martin BC, *et al*. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part III. *Value Health* 2009; **12**(8): 1062–73.

158 Willke RJ, Mullins CD. "Ten commandments" for conducting comparative effectiveness research using

"real-world data." *J Manag Care Pharm* 2011; **17**(9 Suppl A): S10–S15.

159 Dreyer NA, Schneeweiss S, McNeil BJ, *et al.* GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010; **16**(6): 467–71.

160 Luce BR, Drummond MF, Dubois RW, *et al.* Principles for planning and conducting comparative effectiveness research. *J Comp Eff Res* 2012; **1**(5): 431–40.

161 Morton JB, McConeghy R, Heinrich K, *et al.* Consensus of recommendations guiding comparative effectiveness research methods. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1354–60.

162 Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998; **316**(7127): 285.

163 Godwin M, Ruhland L, Casson I, *et al.* Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003; **3**: 28.

164 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005; **365**(9453): 82–93.

165 Devereaux PJ, Bhandari M, Clarke M, *et al.* Need for expertise based randomised controlled trials. *BMJ* 2005; **330**(7482): 88.

166 Ernst E, Canter PH. Limitations of "pragmatic" trials. *Postgrad Med J* 2005; **81**(954): 203.

167 Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009; **10**: 37.

168 Kent DM, Kitsios G. Against pragmatism: on efficacy, effectiveness and the real world. *Trials* 2009; **10**: 48.

169 Eldridge S. Pragmatic trials in primary health care: what, when and how? *Fam Pract* 2010; **27**(6): 591–2.

170 Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011; **13**(2): 217–24.

171 Ware JH, Hamel MB. Pragmatic trials: guides to better patient care? *N Engl J Med* 2011; **364**(18): 1685–7.

172 Mitka M. FDA advisory decision highlights some problems inherent in pragmatic trials. *JAMA* 2011; **306**(17): 1851–2.

173 Chalkidou K, Tunis S, Whicher D, *et al.* The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials* 2012; **9**(4): 436–46.

174 Ratner J, Mullins D, Buesching DP, *et al.* Pragmatic clinical trials: U.S. payers' views on their value. *Am J Manag Care* 2013; **19**(5): e158–e165.

175 Kim SY, Miller FG. Informed consent for pragmatic trials: the integrated consent model. *N Engl J Med* 2014; **370**(8): 769–72.

176 Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials. *JAMA* 2014; **311**(23): 2381–2.

177 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**(4): 409–22.

178 Thorpe KE, Zwarenstein M, Oxman AD, *et al.* A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009; **62**(5): 464–75.

179 Anderson ML, Griffin J, Goldkind SF, *et al.* The Food and Drug Administration and pragmatic clinical trials of marketed medical products. *Clin Trials* 2015; **12**(5): 511–19.

180 Loudon K, Treweek S, Sullivan F, *et al.* The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015; **350**: h2147.

181 Zwarenstein M, Treweek S, Gagnier JJ, *et al.* Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008; **337**: a2390.

182 Carey TS, Melvin CL, Ranney LM. Extracting key messages from systematic reviews. *J Psychiatr Pract* 2008; **14**(Suppl 1): 28–34.

183 Waters E, Armstrong R. Knowledge synthesis, translation and exchange in public

health and health promotion: the role of a Cochrane Review Group. *Promot Educ* 2007; **14**: 34–5.

184 Institute of Medicine. IOM Standards for Developing Trustworthy Clinical Practice Guidelines. February 2011. http://www.nationalacademies.org/hmd/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx (accessed May 2019).

185 Kanouse DE, Kallich JD, Kahan JP. Dissemination of effectiveness and outcomes research. *Health Policy* 1995; **34**: 167–92.

186 Fenton JJ, Weyrich MS, Durbin S, *et al*. Prostate-specific antigen-based screening for prostate cancer: a systematic evidence review for the US Preventive Services Task Force: Evidence synthesis No. 154. AHRQ publication 17-05229-EF-1. Rockville, MD: AHRQ, 2018.

187 Etzioni R, Penson DF, Legler JM, *et al*. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst* 2002; **94**(13): 981–90.

188 Pollack A. The minimal impact of a big hypertension study. *New York Times, Nov* **28**, 2008. https://www.nytimes.com/2008/11/28/business/28govtest.html (accessed May 2019).

189 Timbie JW, Fox DS, Van Busum K, *et al*. Five reasons that many comparative effectiveness studies fail to change patient care and clinical practice. *Health Aff (Millwood)* 2012; **31**(10): 2168–75.

190 Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. *Med Decis Making* 1991; **11**(3): 154–8.

191 Emanuel EJ. What cannot be said on television about health care. *JAMA* 2007; **297**(19): 2131–3.

192 Institute of Medicine. Clinical practice guidelines we can trust. Washington, DC: National Academies Press, 2011. http://www.nationalacademies.org/hmd/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx (accessed May 2019).

193 Mendelson A, Kondo K, Damberg C, *et al*. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Ann Intern Med* 2017; **166**(5): 341–53.

194 Patient-Centered Outcomes Research Institute. Putting evidence to work. https://www.pcori.org/research-results/putting-evidence-work (accessed September 2018).

195 Patient-Centered Outcomes Research Institute. Dissemination and implementation. https://www.pcori.org/sites/default/files/PCORI-Dissemination-Implementation.pdf (accessed September 2018).

196 Gaige M. Comparative effectiveness research: a valuable tool for health-care reform? *Penn Medicine* 2009/2010; **Winter**: 24–31, p. 29.

197 Mitka M. US government kicks off program for comparative effectiveness research. *JAMA* 2010; **304**: 2230–1.

198 Wilensky GR. Cost-effectiveness information: yes, it's important, but keep it separate, please! *Ann Intern Med* 2008; **148**: 967–8.

199 Centers for Medicare & Medicaid Services. Projected national health expenditure data. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected.html (accessed September 2018).

200 Hochman M, McCormick D. Characteristics of published comparative effectiveness studies of medications. *JAMA* 2010; **303**: 951–8.

201 Bourgeois FT, Murthy S, Mandl KD. Comparative effectiveness research: an empirical study of trials registered in ClinicalTrials.gov. *PLoS One* 2012; **7**(1): e28820.

202 CORDIS. Comparative Effectiveness research on Psychiatric HOSpitalisation by record LINKage of large administrative data sets. https://cordis.europa.eu/project/rcn/185457/factsheet/en (accessed May 2019).

203 Health and Medical Research Fund. About HMRF. https://rfs1.fhb.gov.hk/english/funds/funds_hmrf/funds_hmrf_abt/funds_hmrf_abt.html (accessed October 2018).

204 Sherman RE, Anderson SA, Dal Pan GJ, *et al*. Real-world evidence: what is it and what can it tell us? *N Engl J Med* 2016; **375**(23): 2293–7.

205 US Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and Food and Drug Administration Staff. 2017. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM513027.pdf (accessed November 2018).

206 US Food and Drug Administration. Real world evidence. https://www.fda.gov/scienceresearch/specialtopics/realworldevidence/default.htm (accessed November 2018).

207 Plueschke K, McGettigan P, Pacurariu A, *et al*. EU-funded initiatives for real world evidence: descriptive analysis of their characteristics and relevance for regulatory decision-making. *BMJ Open* 2018; **8**(6): e021864.

208 Pfizer Inc. BMS-Pfizer alliance to unveil real-world data analyses – cost, safety and comparative effectiveness findings associated with oral anticoagulants in non-valvular atrial fibrillation. 2017. https://investors.pfizer.com/investor-news/press-release-details/2017/BMS-Pfizer-Alliance-to-Unveil-Real-World-Data-Analyses---Cost-Safety-and-Comparative-Effectiveness-Findings-Associated-with-Oral-Anticoagulants-in-Non-Valvular-Atrial-Fibrillation/default.aspx (accessed November 2018).

209 Tepper PG, Mardekian J, Masseria C, *et al*. Real-world comparison of bleeding risks among non-valvular atrial fibrillation patients prescribed apixaban, dabigatran, or rivaroxaban. *PLoS One* 2018; **13**(11): e0205989.

210 Flacco ME, Manzoli L, Boccia S, *et al*. Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *J Clin Epidemiol* 2015; **68**(7): 811–20.

211 European Commission. Medicines for children: major developments. https://ec.europa.eu/health/human-use/paediatric-medicines/developments_en (accessed October 2018).

212 Murray MD. Curricular considerations for pharmaceutical comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2011; **20**(8): 797–804.

213 Marie Skłodowska-Curie Actions. Research fellowship programme. https://ec.europa.eu/research/mariecurieactions (accessed October 2018).

214 National Institute for Health Research. NIHR fellowships. https://www.nihr.ac.uk/funding-and-support/funding-for-training-and-career-development/fellowship-programme.htm (accessed October 2018).

215 Osler W. *Aequanimitas: With Other Addresses to Medical Students, Nurses and Practitioners of Medicine*. Birmingham, AL: Classics of Medicine Library, 1987. Originally published 1904.

216 Egnew TR. The art of medicine: seven skills that promote mastery. *Fam Pract Manag* 2014; **21**(4): 25–30.

217 Kerridge I, Lowe M, Henry H. Ethics and evidence based medicine. *BMJ* 1998; **316**: 1151–3.

218 Greenhalgh T, Howick J, Maskrey N, *et al*. Evidence based medicine: a movement in crisis? *BMJ* 2014; **348**: g3725.

27

Data Mining and Other Informatics Approaches to Pharmacoepidemiology

Andrew Bate[1,2], Gianluca Trifirò[3], Paul Avillach[4], and Stephen J.W. Evans[5]

[1] *Pfizer Ltd, Walton-on-the-Hill, Surrey, UK*
[2] *Department of Medicine, NYU School of Medicine, New York, NY, USA*
[3] *Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of Messina, Messina, Italy*
[4] *Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA*
[5] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK*

Recently our capability to harness computational power to collect, share, and analyze data has increased dramatically year on year. This has led to vast data sources potentially able to provide answers to questions relevant to pharmacoepidemiology. At the same time, there remains a need for method and tool development to extract reliable insights, rather than simply more information, distinguishing signals from the noise inherent in large volumes of data that are not collected with research as their primary purpose. Effective and reliable drug- and vaccine-related evidence generation is the goal in order to improve the public's health. Informatics and data-mining approaches are not intended to replace traditional pharmacoepidemiologic approaches, but rather to enhance the field by addressing questions of interest more efficiently, and being able to answer additional ones. This chapter sets out to provide an overview of how data mining and other informatics approaches are applied to the field of pharmacoepidemiology.

Data mining should be seen as exploratory data analysis for hypothesis generation, and as part of a knowledge discovery process [1]. It looks to uncover patterns or correlations in the dataset with no or limited presupposition, but almost invariably requires more rigorous testing of any emerging hypothesis, tailoring the subsequent testing to the issue at hand. As some researchers have put it, "data mining is asking a processing engine to show answers to questions we do not know how to ask" [2]. Often in pharmacoepidemiology "data mining" is used synonymously with quantitative signal detection, although naturally other forms of hypothesis generation of relevance to pharmacoepidemiology would also be in scope. As the sources of existing healthcare data and our capability to access and potentially analyze them increase, we strive to maximize the value of the data through developing more effective data mining and processes associated with it. Hence the recent emphasis on "Big Data" and "real-world evidence" and

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

the supposed capability to generate more rapid insights from the increasing amount of diverse data available (see, for example, [3–6]). However, the wide range of different terminologies, dictionaries, and controlled vocabularies that are used as well as nonrandom variability in data capture due to different healthcare systems make linkage and analysis of multiple data sources challenging. In addition, there is an increasing amount of unstructured free-text/narrative data recorded, as well as repositories with biologic data, chemical structure, and pharmacogenomics information. However, data protection laws may make access to text and linkage between data sources more difficult in many regions. Informatics approaches, defined as "the study of the structure, algorithms, behaviour, and interactions of natural and artificial computational systems" [7], have obvious value for data mining, while they are also of relevance in the wider application of pharmacoepidemiologic approaches.

Signal detection, refinement, and evaluation describe three distinct secondary uses of data [8] that together form a signal management continuum (Figure 27.1). There is much discussion of definitions of signal-related terms and their relations as components of safety surveillance [9]. Other terms that are used include "signal strengthening" in place of "signal refinement," "signal validation" in place of "signal evaluation," and "signal substantiation," which incorporates an element of both refinement and evaluation, but has been applied using external data, such as protein information, to provide a pharmacologic mechanism for detected signals, often through semi-automated data linkage and analysis [6,10,11]. There is some inconsistency in the use of signal-related terms and further terminologic harmonization would be beneficial.

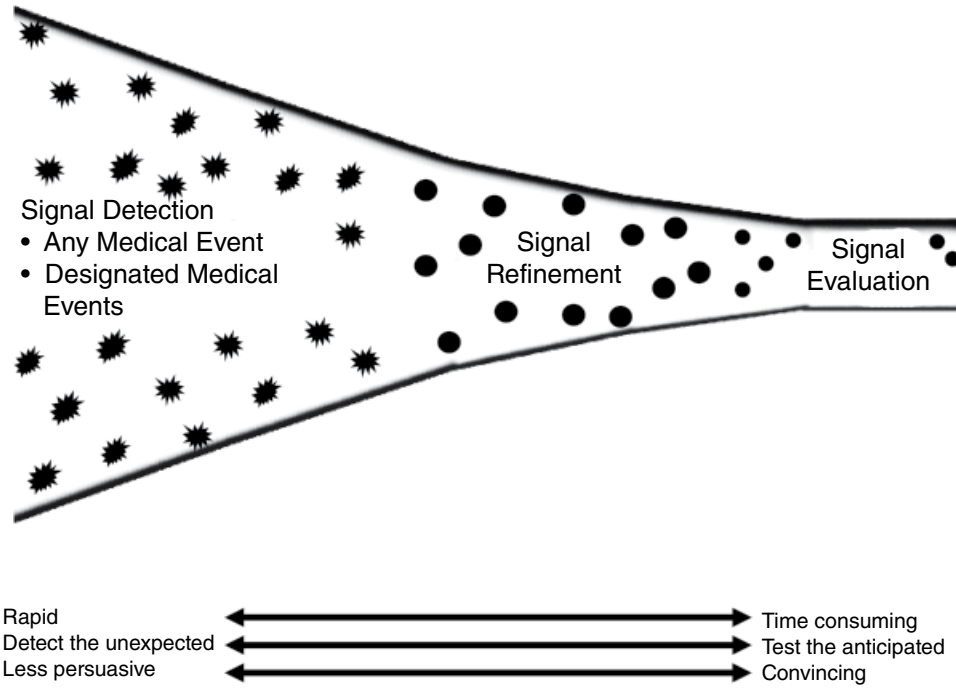


Figure 27.1 Signal stages after product approval and launch.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Signal Detection and Data Mining: Realistic Expectations

All medicines are licensed on the basis of a favorable benefit/risk balance. Routine activities are aimed at supporting the continual evaluation of the benefit/risk profile of medicines in the postmarketing setting. Specific aims include primarily detection of drug safety signals, and then quantification, characterization, and identification of predictors of known adverse drug reactions (ADRs).

The concept of a signal, from a drug or vaccine surveillance point of view, has evolved from its definition by the World Health Organization (WHO) as "reported information on a possible causal relationship between an adverse event (AE) and a drug, the relationship being unknown or incompletely documented previously. Usually more than a single report is required to generate a signal, depending upon the seriousness of the event and the quality of the information" [12], to a more comprehensive definition according to the Council for International Organizations of Medical Sciences (CIOMS): "information arising from one or multiple sources (including observations or experiments), which suggests a somehow new, potentially causal association between administration of a medicine and adverse event(s) that is of sufficient likelihood to justify verificatory action" [13].

In particular, this new definition highlights that a signal can be detected from multiple data sources. In a review of 25 drugs withdrawn from the market for safety reasons from 2000 to 2010 in the US or European Union (EU), the type of data that provided the basis for each withdrawal was examined [14]. The majority of safety-based withdrawals were rare events such as progressive multifocal leukoencephalopathy, Stevens–Johnson syndrome and cardiac arrhythmias, or AEs with delayed onset, such as myocardial infarction and liver toxicity. In almost 50% of the drug withdrawals, the authors' review [14] suggested that spontaneous reporting systems (SRSs, see Chapter 10) were the primary source of data in triggering regulatory action, thus confirming that SRSs remain the cornerstone of pharmacovigilance. In contrast, in two instances (8%) for two outcomes – cardiovascular events (including myocardial infarction and stroke) and liver toxicity/veno-occlusive disease – randomized clinical trials (RCTs) were the sole source of the safety information. For the remaining 48% of withdrawals, a combination of data from SRS, clinical trials, and/or observational studies contributed to the regulatory action.

Safety signal detection is both an iterative and a dynamic process, since emerging safety issues can be encountered over time after approval (e.g., progressive multifocal leukoencephalopathy associated with natalizumab). ADRs may have many different types of manifestation, underlying mechanisms, frequencies, latency, and predictors (see Chapter 3), which may require integrating and understanding evidence from all possibly relevant information sources on drug safety. Continued development of multimodal signal detection requires a deeper understanding of the data sources used and further research on methods to generate and synthesize signals [15].

Although RCTs are considered to be the most rigorous approach to determining a cause-and-effect relationship between an intervention (e.g., medication exposure) and an outcome, these trials are generally designed and powered to assess efficacy rather than safety end-points. The controlled nature of such trials, however, calls for a limited number of patients who may be at a lower baseline risk of ADRs than the population of all potential users of the drug, so the statistical power to detect ADRs will be low (see Chapter 4). In addition to limited study sample size, a relatively short observation period makes it difficult to detect ADRs that have long

latency [16,17]. Similar ADRs that occur in vulnerable subpopulations or when medications are in practice used suboptimally or in unusual combinations are also particularly important to detect once medicinal products are utilized in routine healthcare. While RCTs in the premarketing phase and SRSs in the postmarketing setting remain essential for drug safety surveillance, there are still gaps that may be filled by observational data derived from different sources. In particular, the role of mining data from healthcare databases for signal detection has been more extensively investigated in recent years [6]. Compared to clinical trial data, population-based healthcare databases such as electronic health records and claims databases (see Chapters 11–14) contain data from clinical practice about larger populations with longer follow-up periods, which may help in the detection of adverse reactions occurring after long-term exposure to medicines (e.g., cardiac valve disorders). Preliminary results [18,19] showed that these data sources might complement SRSs in routine pharmacovigilance in the postmarketing setting for detecting specifically signals concerning events occurring at relatively high frequency in the general population, as well as those that are multifactorial (e.g., myocardial infarction) and as such not commonly considered as a potentially drug-induced event, so unlikely to be reported to any SRS.

Ideal pharmacovigilance systems for signal detection should allow timely and accurate identification of new, potentially causal associations of drug(s) and AE(s) requiring further signal management, while minimizing false positive signals and optimizing use of human resources. Once detected, in general, a suspected signal has to be strengthened (or weakened) by adding other evidence on biologic substantiation and other key information on causality assessment, as done experimentally in the EU-ADR project (Figure 27.2) [10], and finally validated and fully characterized through formal pharmacoepidemiology studies. Because of the substantial effort required to evaluate signals, detecting these in systems with low precision (high false positive rates) would not be sustainable routinely. Ongoing efforts to develop quantitative methods, or other approaches that can provide at least some semi-automated triage, all with strong performance characteristics (positive and negative predictive values) in highlighting potential signals, remain critical. While improving signal detection is important, and is the focus of much of this chapter, our capability to effectively rule out spurious signals, and rapidly and effectively further analyse and better understand true emerging risks, through signal refinement and evaluation activities is equally critical, and is addressed in other chapters.

Disaster-Driven Pharmacovigilance

The traditional pharmacovigilance system mostly based on SRSs has been in place since the end of the 1960s as a reaction to the tragedy of thalidomide-induced limb defects. This system, despite its great potential, is far from optimal as a consequence of well-known SRS limitations such as underreporting, notoriety bias, and lack of denominators. It is therefore not surprising that serious ADRs leading ultimately to drug withdrawal are sometimes detected only with significant delay.

Electronic health record (EHR) data may be able to identify new risks for drugs associated with AEs that have high background incidence rates (such as acute myocardial infarction), as well as events that are not pharmacologically predictable and are less likely to be suspected as drug induced, thus less likely to be reported and therefore in spontaneous reporting systems. Historically, spontaneous reporting systems have been good at highlighting the unpredictable (at the time), for instance coughing with angiotensin-converting enzyme (ACE) inhibitors, heart rhythm disorders with terfenadine, or progressive multifocal leukoencephalopathy

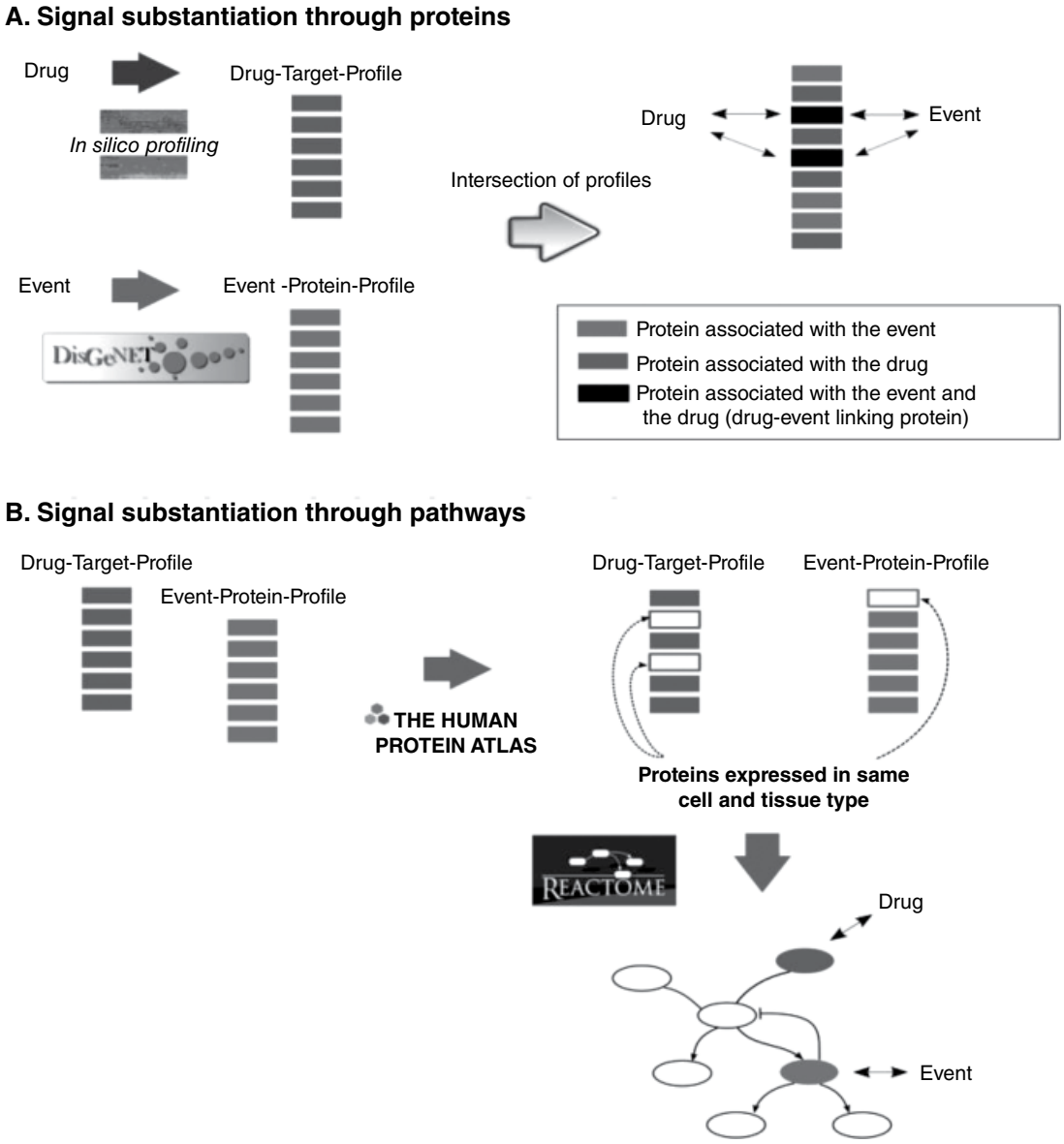


Figure 27.2 Example of a process for signal substantiation through proteins. *Source:* Bauer-Mehren A, van Mullingen EM, Avillach P, *et al*. Automatic filtering and substantiation of drug safety signals. *PLoS Comput Biol* 2012; **8**(4): e1002457.

(PML) and natalizumab and heart valve disorders with fenphen (the combination product of fenfluramine and phentermine). Data from EHRs provide denominators for the rate of occurrence as well as greater detail regarding possible confounding factors such as patient demographics, drug use, and utilization of healthcare services. These data may permit

epidemiologic evaluation of the benefit/risk profile of drugs (see Chapter 35). This process then puts safety issues in a broader perspective and fosters sound regulatory decisions, therefore providing a data source for moving from signal detection across signal refinement toward signal evaluation. This can lead to a more compelling rationale for further investigation than might sometimes be the case with signals from SRSs with limited clinical plausibility at the time.

The Example of Rofecoxib's Withdrawal from the Market and the Potential Role of Healthcare Databases

Several initiatives have been exploring the role of healthcare databases for drug safety signal detection and signal management. An example that has been studied in depth is the case of withdrawal from the market of rofecoxib due to increased risk of cardiovascular thrombotic events. Rofecoxib is a prescription cyclooxygenase (COX)-2 selective nonsteroidal anti-inflammatory drug (NSAID) previously indicated for relief of osteoarthritis signs and symptoms, management of acute pain in adults, and treatment of menstrual pain. The European Medicines Agency (EMA) and the United States Food and Drug Administration (FDA) provided approval for rofecoxib in 1999. In June 2000, the results of the Vioxx Gastrointestinal Outcome studies (VIGOR) trial were submitted to the FDA and demonstrated an increased risk of cardiovascular thrombotic events. This trial showed that patients who were given rofecoxib had four times as many myocardial infarctions as those who were given naproxen, which was attributed by the authors of the VIGOR trial at that time as due solely to the cardioprotective effect of naproxen. Subsequently, in September 2004, the Adenomatous Polyp Prevention on Vioxx (APPROVe) study showed increased risk of myocardial infarction and stroke with rofecoxib as compared to placebo after 18 months of treatment. The same month, the manufacturer withdrew rofecoxib from the market because of concerns about increased risk of heart attack and stroke associated with long-term, high-dosage use [20]. Thus, it took almost five years for rofecoxib to be withdrawn from the market after the first evidence of increased myocardial infarction risk was documented [21]. By the time of rofecoxib's withdrawal, more than 100 million prescriptions had been filled in the US, with tens of millions of these prescriptions being written for persons who had a low or very low risk of gastrointestinal complications of nonselective NSAIDs [20], which was the rationale for developing COX-2 selective NSAIDs. Using actual penetration of rofecoxib in the market, it was calculated that if the medical records of 100 million patients had been available for safety monitoring, the adverse cardiovascular effect would have been discovered in just three months [22]. In the context of the FDA's Sentinel initiative, a modular program was built that semi-automatically performs cohort identification, confounding adjustment, aggregation, and effect estimation across multiple databases, and application of a sequential alerting algorithm (see Chapter 25). In retrospective assessments, the system identified an increased risk of myocardial infarction with rofecoxib and an increased risk of rhabdomyolysis with cerivastatin years before these drugs were withdrawn from the market [23]. Similarly, the Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge (EU-ADR) project, which built a European multidatabase network covering more than 30 million citizens, has shown that, had this data source been available at the time rofecoxib was still marketed, there could have been a faster withdrawal [18]. Although international SRSs did not play a significant role in detection and prioritization of this emerging signal, it is worth noting that the Dutch SRS did identify this as a potential signal [24].

What are the lessons to be learned from this story? First, suggestions to replace spontaneous reporting by other systems are illogical and not supported by the evidence provided by the rofecoxib case, as signals of rofecoxib-associated

cardiotoxicity from both clinical trial and Dutch SRS were detected; secondly, it is what happens after the first signal of an ADR has been found that should be the major focus of concern. In this case there was a considerable delay from the initial signal to regulatory action [24]. Availability of a large amount of EHR data represents an additional data source for signal refinement, as well as for signal evaluation, thus potentially contributing to accelerating the process of signal management. Nevertheless, as discussed shortly, while healthcare databases on their own or combined in networks are used for signal refinement, as discussed in Chapter 25, they may also help enhance signal detection capability.

As another example, in recent years safety data from the US Vaccine Adverse Event Reporting System (VAERS) and the Vaccine Safety Datalink suggested that more venous thromboembolism (VTE) cases were observed than expected after vaccination with quadrivalent human papillomavirus vaccine (HPV4; Gardasil®), but data were inconclusive [25,26]. The Sentinel initiative rapidly evaluated the risk of VTE in more than 650 000 females aged 9 through 26 years of age, totalling more than 1.4 million doses of Gardasil evaluated, and found no evidence of an increased risk for VTE after Gardasil vaccination, reporting a relative risk estimate (95% confidence interval) from both unrestricted analyses, for all doses, with a 28-day risk interval of 0.7 (0.3–1.4) [27].

Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

Data Mining and Causality

The fundamental problem for much of pharmacoepidemiology is that of inferring causality. For much data mining outside of pharmacovigilance, there is no need to assess causality: association is enough. This can be true even for public health purposes in some circumstances. For example, let us assume that there is no biologic or pharmacologic effect of isotretinoin on the likelihood of a patient's taking a suicidal action. However, there may well be an association resulting from other conditions the patient has, which result in their being prescribed isotretinoin. From a public health perspective, it would be good for the prescribing health professional to be aware of the isotretinoin–suicide association, even if it is not causal. That is, when prescribing isotretinoin, the characteristics of the patient and the circumstances in which it is prescribed are associated with suicidal events, and attention should be paid to caring for the patient with this in mind. With data mining in the sphere of economics or politics, the association between previous purchases and other factors like age would be enough to make the targeting of an advertisement to an individual very productive, in that the individual is more likely than others to respond to that particular advertisement, without any requirement to take other factors into account that may be the causal elements. A similar effect is seen in response to encouraging donations or going out to vote for a particular candidate; no inferences about what is causal are necessary.

Finding associations is a reasonable first step toward the assessment of causality. Associations that are stronger may be believed to be good candidates for both being causal and of public health importance. However, Ioannidis has argued, persuasively, that almost all of Bradford Hill's considerations (not "criteria," as Ioannidis and others have misquoted) for causality have limitations, and maintains that *strength* of association is almost the exact opposite of a good indicator – "when large effects are seen, they are mostly transient and almost always represent biases and errors" [28]. Ioannidis's assessment is largely based on randomized trials, but those are where evidence for causality is likely to be good. He notes that this applies to adverse effects as well as benefits, but has not assessed it

in relation to data mining. That said, it seems likely the same arguments could apply.

Many of the large associations, whether seen in pharmacovigilance, randomized trials, or genome studies, are based on small numbers of exposed cases. Therefore, careful attention to lower limits of confidence intervals may help with considering the magnitude of association. In data mining for pharmacovigilance, it is often the case that cut-offs for signals are based on lower limits of confidence Intervals, and this should probably be a principle in data mining more generally. It may even be argued that it is sensible to use 99%, 99.5%, or even more extreme confidence limits to account for multiplicity [29]. This may lead to too many type II statistical errors, but the problem with method validation in pharmacovigilance is that many of the labeled "true positive" associations could well be false positives themselves, so what is assumed to be a "gold standard" is in fact an "iron pyrites standard" (iron pyrites is known as "fool's gold"!). In addition, Bayesian shrinkage estimates also reduce the large effects seen with small sample sizes. The "shrinkage" can be illustrated using a comparison of an observed count of the numbers of an AE compared with the expected count for that event. The expected count in pharmacovigilance may be derived from a database of spontaneous reports, or from a comparison group in a trial or a pharmacoepidemiologic study. Most measures of association are based on the ratio of the observed (o) with the expected (e) count. A shrinkage estimate will add a constant, usually 0.5, to both the observed and the expected count, and recompute their ratio. If the observed count is, say, 2 and the expected is only 0.1, then o/e = 2/0.1 = 20. This ratio (close to a risk ratio or an odds ratio) is quite extreme, but using (o+0.5)/(e+0.5) gives 2.5/0.6 = 4.2. This is a much less extreme value and the appropriate lower confidence interval will also be much lower. Values are "shrunk" toward the null value of 1. For a more extensive treatment of shrinkage for signal detection, see [30]. The ranking of potential signals must take these considerations into account, so that large (relative) effects with very small observed numbers of cases do not dominate. At the very least, the usual cut-off criteria utilize lower bounds of confidence (or in Bayesian systems, credible) intervals.

There is an extensive literature on how potential signals are highlighted and investigated as part of sophisticated auditable systems for signal management [31–33]. Such management systems include alternative approaches for highlighting emerging signals from spontaneous reports and other sources; this is beyond the scope of this chapter, but suffice it to say that quantitative signal detection plays a core role as the spontaneous reporting repositories increase in size. The whole process is in contrast to *data dredging*, where a search is made for one or a few statistically significant findings, without reporting that such a search has been made and without taking into account the problems of multiple testing. With data mining, it is explicit that many possible associations are found, and further processes will be carried out to find those that are likely to be causal. The wider knowledge discovery process for using data mining as a component of a signal management system has been treated extensively in the pharmacovigilance literature, for example [34, 35].

The "COMPARE" project of Ben Goldacre and colleagues [36] has established that many investigators change the outcome variable specified in the protocol (presumably to get a desirable answer), and this has been found by many others, in the context of both trials [37] and systematic reviews [38]. With transparent methods of data mining, the output should not be treated as if it were a confirmed hypothesis. This is at least partly because bias and confounding are especially challenging for data mining and signal detection. In epidemiologic studies where specific hypotheses are tested, it is possible to consider biases and confounding relating to a specific drug/AE combination. In data mining it

is much more difficult to deal with biases or confounding in a general way. At the very least it can require massive computing power to be able to carry out, for example, high-dimensional propensity score methods, calculating a new score for every one of possibly many thousands of drug/event pairs, and then using a regression method to look at each association.

Hence data mining, which finds associations that may or may not be causal, can nevertheless be useful in some contexts, provided the associations are not taken as providing sufficient evidence on their own of causal effects [39]. Also, methods for association detection can be used to provide ranking for further triage, as well as more granular hypotheses that can be prioritized for future hypothesis-testing studies and helping to refine hypothesis definitions. There is a necessarily iterative element, in that generated or tested hypotheses may stimulate the need for further exploratory analysis, and the process will continue. This broad conceptual use of data mining/signal detection makes rigorous assessment of performance and value challenging. Despite this openness, the credibility and value of data mining and exploratory data analysis (EDA) are supported by clear process and audit trails for the use of such techniques, as exhibited for example in signal detection in SRS as part of overall signal management plans [31].

Using Replication and Transparency to Improve Reliability

Ioannidis has argued that replication is an important component of improving the quality of published findings [40]. This also applies when data mining. In this case, the combining of results across different settings and databases helps to distinguish genuine causal effects from spurious associations. Replication can occur both between different independent groups studying the same question(s) and also within a single group or collaboration. If data mining is to be done across different groups, then a great deal more transparency is required. The Sentinel initiative has an excellent record of doing this, making code lists and software readily available, as does the group called Observational Health Data Sciences and Informatics (OHDSI, pronounced "Odyssey") and its predecessor, the Observational Medical Outcomes Partnership (OMOP). OHDSI includes among its objectives "Reproducibility: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement." This can be a deliberate choice and is open to other investigators to follow its example, although there are many challenges with trying to ensure reproducibility and transparency.

However, the transparency did allow Gruber *et al.* [41] to examine a particular result from the OMOP analyses that they felt was anomalous. The resources required to reexamine more than one such result were noted to be considerable. This anomalous result, a finding of upper gastrointestinal bleeding associated with benzodiazepines, was studied in detail. The conclusion was that the association was largely restricted to the first day of a prescription and unlikely to be a causal finding – "driven by an excess of procedures on the first day of treatment." This seems entirely sensible as an explanation, since benzodiazepines are given prior to endoscopy. Gruber *et al.* conclude: "It is likely that all surveillance programs will need tailored designs that reflect pharmacologic and clinical knowledge." To some degree, this misses the point: if each possible association requires an individual design, it suggests that surveillance is impossible across more than a limited range of issues. What is clear is that, as stated earlier, taking an association that is flagged up by a system as if it were a confirmatory study is a mistake. Further work is required, as in the processes of signal detection leading to confirmation (or not) through signal evaluation. It may be more sensible to ensure that the window during which a possible risk may appear should not generally include the day of a prescription.

Guidelines like the International Society for Pharmacoepidemiology–International Society for Pharmacoeconomics and Outcomes Research (ISPE-ISPOR) guidance on reporting of healthcare database studies [42] and the guideline for the REporting of studies Conducted using Observational Routinely collected Data (RECORD) statement [43] from the "Enhancing the QUAlity and Transparency Of health Research" (EQUATOR) network (http://www.equator-network.org), which includes the CONsolidated Standards Of Reporting Trials (CONSORT) guidelines, may help with this.

Problems of Exploratory and Confirmatory Analyses

Data mining in surveillance of medicines and similar fields is exploratory in nature, as already noted, and generally takes place in one or a few databases. When confirmation of "signals" or other hypotheses that are generated is sought, then it is not generally sensible to carry this out in the same dataset. It is not good enough simply to divide a single dataset (randomly) into training and test sets. A random split will nearly always ensure that the test set will replicate the biases in the training set. It only deals with sampling variability, and provides some limited protection when the model used in the data mining has included factors that are there just because of chance variation. The test set will not reflect the chance factors, but will reproduce the biases. This is why many predictive models developed in one context do not function as well as expected when they are applied in different data. Nonrandom splitting (e.g., by region) may be possible, but totally independent data are required for proper validation.

In practice, however, with the finite, even if large, number of observational databases in total, many will be unsuitable (in terms of variables recorded, data access, and power) to answer some questions. For example, for some questions smoking and lifestyle factors may be vital to address confounding and the number of appropriate independent databases available may be very limited. Frequently there are different studies including both exploratory and confirmatory analyses, which are commonly conducted in the same or overlapping databases. It is not clear that totally independent analyses can be carried out in the same databases, but reanalysis by independent groups may come to different conclusions, showing that there is more than sampling uncertainty (captured in a confidence interval) in the findings. Walker [44] has suggested that it is possible to use the same database and employs the term "orthogonality" to describe assessing other hypotheses related to a primary discovery, which nevertheless add to or subtract from the evidence for causality of the primary discovery. He uses an example of intussusception following a rotavirus vaccine (the primary hypothesis, found by "data mining") and utilizing the same data to see if the vaccine is associated with reports of other conditions that have the same pathology. While Walker argues that this is evidence for causality, Gould [45] disagrees, stating that "hypothesis-generating and test data from the same data source generally cannot be considered 'independent,'" and arguing that the evidence for causality may not be strengthened at all by this approach if there is unmeasured or unanalyzed confounding. Confounding by indication can also apply across related pathologies. Whatever the view, it is important to be clear that overinterpreting concordance with previous findings as if it were as convincing as two completely independent studies is likely to be mistaken. Using the "orthogonal" hypothesis may serve to strengthen evidence, but will rarely be sufficient to confirm a causal relationship.

Fortunately, databases themselves often have a relatively rapid turnover of included patients (e.g., US claims databases), so there can be a reduction in some of the dependence when conducting studies in the same datasets with different time periods. Clear transparency of

exactly how outcomes, exposures, and confounding variables are measured, classified, and utilized in a statistical model is necessary. As always in science, findings need to be related to previous relevant findings to ensure appropriate interpretation by the reader of a single study publication. When exploratory and confirmatory analyses are executed on the same drug safety issue in the same database, it will be important to know what data influenced the design of the confirmatory study and how, and the extent to which the authors (or readers) believe this influenced the results.

The phrase "signal refinement" may be applied to a process akin to Walker's orthogonal approach, in that outcomes related to the primary hypothesis will also be examined. Similarly, looking at whether other related exposures (perhaps the same drug class) have the same effect and whether the pattern with time is consistent contribute importantly to the evaluation. This process of signal refinement or rapid cycle analyses [46] looks more like full studies, and often leads to a full hypothesis testing confirmatory study. In Walker's example [44], having found an association with intussusception, the search for similar associations with gastrointestinal hemorrhage, intestinal obstruction, gastroenteritis, and abdominal pain makes such an exploratory search begin to look like a more formal study. They are exploratory in mindset but with a clear hypothesis on the table. Although the approach is more common now, it has been used previously in the literature, for example by Behrens [47] as a "rough confirmatory analysis." As the boundary between EDA and confirmatory data analysis (CDA) becomes blurred, it will be important that transparent processes are in place to ensure appropriate credible inferences are made from analyses.

A challenge of particular importance in data mining and hypothesis generation is the necessity of rapidity – and the delays in data capture in data sources and their availability for secondary data analysis for pharmacoepidemiology can be a key rate limiter. Developing capabilities for analysis as data, and confidence in its accuracy, accrue will be very important, and clearly the balance between availability and ensuring robust quality is a natural tension. Transparency around process and near real-time quality control of the raw data to the extent possible is required, together with the capability to update analyses over time as needed.

Currently Available Solutions

Data Mining in Spontaneous Reports

The history of routine data mining in pharmacoepidemiology began with the analysis of spontaneous reports of suspected adverse drug reactions [48–51]. It was initially very simple, based only on knowledge from a database of suspected drugs and AEs reported with them. For every combination of drug (classified as drug substance) and AE (classified at, usually, preferred term level according to a hierarchical terminology), it produced a measure of association (disproportionality), then used a cut-off for that measure to classify whether this combination results in a "signal of disproportionality" or not (and subsequent clinical review, leading some of these to be "signals of suspected causality"). The core concepts had been proposed in fact many years previously, though the computing technology had not been applied to every possible combination in a database [52]. Specifically, measures of disproportionality use a two-by-two contingency table for each drug–event pair and look for unexpectedly frequent reporting of the combination compared to the count expected based on general reporting of the drug and the AE if they were independent. The four most commonly used metrics – that is, information component (IC), empirical Bayes geometric mean (EBGM), proportional reporting ratio (PRR), and reporting odds ratio

(ROR) – are all based on this core contingency table [35]. Despite the inherent limitations and oversimplification of the complexities of spontaneous reporting, these metrics have been found to have robust performance characteristics in practice, in the sense that the precision and recall estimates from many independent method evaluations are concordant in suggesting they provide helpful assessments of the data (see, e.g., [53,54]). Thus, they are now widely considered sufficiently strong to justify the usefulness of such approaches as routine surveillance tools for signal detection by regulators and industry [35].

Data mining has looked to include other factors available for individual reports – age, gender of patients, year of reporting, and country in which it was reported [55] – which would modify the metric used. Approaches such as lasso shrinkage regression, a Bayesian approach applicable for regression across the often huge numbers of explanatory variables needed for hypothesis generation, can be used to analyze all possible medicinal products and combinations thereof with a given AE as predictor variable [56]. In contrast to traditional regression approaches, lasso shrinkage regression addresses problems with lack of convergence, prohibitive computational complexity, and unstable parameter estimates that can occur with large numbers of predictor variables. This regression approach does not simply look at a single drug–AE combination, but considers what concomitant medication appears on an individual report. So, if two drugs are reported together, it attempts to obtain the separate effect of each. In practice, this added complexity of algorithm and its implementation do not necessarily lead to more effective signal detection, and care is required with implementation [57]. There are other features on such reports that might be utilized in quantitative analysis, for instance previous medical history, concomitant drugs, weight and height, and time since drug initiation. Usually adding features can improve causality assessment in individual cases, as there is more information to consider in the clinical review; however, where the data are frequently missing (and this is nonrandom), it is not guaranteed that consideration of such features in the quantitative screening leads to gains [55,57].

What has been clear in the past is that there are other issues than simply the counts of reports that should determine whether a signal exists or not [35,58]. It has also been shown that each of the systems basically behaves very similarly (that is hardly surprising when they are based on the same fundamental data), and by choice of the cut-off point for a metric can produce very similar results [50]. What is relatively unusual in terms of pharmacoepidemiologic applications is that all the methods applied to spontaneous reports have considered all possible drugs (with thousands of drug substances) and all possible medical terms (again, many thousands of terms). Some methods were initially applied to a narrow selections of drugs or terms, such as two early publications that conducted disproportionality analyses of specific drug–event issues in cefaclor and serum sickness-like symptoms, and hypoglycemia and ACE inhibitors [59,60]. However, to be useful to regulators and those who use the databases of reports, having a system that can consider every possible combination of drug and AE has been important. Research continues to further optimize, from a statistical point of view, quantitative screening approaches in SRS, combined with more work to integrate the medical aspects into signal management systems [61–72].

Data Mining in Healthcare Databases and Distributed Database Networks

A new strand of research looks to use a different source for signal detection: the utilization of healthcare databases, either medical claims data or electronic health records. These databases have been employed for formal pharmacoepidemiologic studies for many years, but using them

for scanning for a huge number of possible associations is more recent. In these databases the variety of data items is several orders of magnitude greater than with spontaneous reports, and data are clearly collected for a very different purpose than signal detection, so there is very rarely any recorded clinical suspicion of adverse drug reactions. Systematic collection on diagnostic and procedural events prior to exposure, in contrast to spontaneous report, presents a significant potential benefit for signal detection. Signal detection and data mining have been tested in both individual databases and also across networks of multiple healthcare databases. The performance of the whole systems for detecting signals of ADRs have not as yet been as great a step forward as many of us anticipated they would be, in that there is no clear emergence of signals that are being detected earlier in healthcare databases than spontaneous reports, and high false positive rates can still occur [19,73,74].

No single observational data source provides a comprehensive view of all encounters a patient accumulates while receiving healthcare, and therefore none can be sufficient to meet all expected outcome analysis needs. This explains the need for assessing and analyzing multiple data sources concurrently. For some research in pharmacoepidemiology, a common data model (CDM) has been used, which is a standardized database model that aims to standardize terminologies of medical events and procedures, and data structures, to facilitate analyses across data sources. This can help with analyzing data across databases and even across countries, but has the disadvantage of necessarily not making full use of all original data and therefore potentially relevant covariates available in some, but not all, databases. The rapidity at which CDM-based analyses can be conducted is particularly advantageous for signal detection/data-mining activities. The alternative is to have optimal designs in a series of studies in different databases, but with enough commonality in the outcome being studied to allow for meta-analysis to provide greater understanding and precision in the answers. Causality may still be difficult to assess [75], but taking a wide picture may still allow for causal conclusions to be reached in some circumstances.

The (in)famous example of hormone therapy (HRT) is instructive. In spite of observational studies suggesting prevention of coronary heart disease (CHD), regulators generally did not allow prevention of CHD as a labeled indication, but did have sufficient belief in the observational studies to include in product information the harms of breast cancer (BC) and VTE. The later randomized trials did not find benefit for CHD, but instead possible harm and did find harms of BC (for combined HRT) and VTE [76]. The lessons of this should be borne in mind when using data mining in observational data [77]. If data mining is seen as a signal detection process, then the failure to eliminate noncausal associations may not be a major problem, provided that the false positive rate is not so high as to overwhelm the systems for evaluation. However, there can be a temptation for regulators to rely on the associations to amend product information, and it clearly would result in better performance for detecting real effects if noncausal associations can be reduced as far as possible.

One issue that has not been well discussed is whether answers from different data sources should automatically be analyzed simply by adding up the numbers across all studies, ignoring the study source ("crude pooling"), or reported separately and combined using formal meta-analysis techniques. It is well known that, even with randomized trials, crude pooling of data can lead to misleading estimates [78]. Since it is misleading with randomized data, it certainly suggests that crude pooling of results should be avoided. Even if there is no difference between the crude pooling and a meta-analysis, the opportunity to test whether results are similar across data sources is important. Whether there is consistency or not in the estimates of

the effect of the drug from different sources can help decide whether an effect is real or not. The consequence is that it is possible that, even with a CDM, using data mining across multiple databases could lead to spurious associations or a failure to find real associations, as occurred in Xu *et al.*'s study, where discordant results were obtained across the OMOP CDM and Mini-Sentinel CDMs and their associated CDM specific analytics tool implementations, all run on the same database [79].

Much of the work done in health records has not attempted to screen for every possible drug and every possible AE, but has focused on a single or a limited number of drug(s) and AE(s), although this is changing. Scaling up the numbers to consider all possible drugs and AEs is not simply computationally challenging, but even with a low false positive rate can mean an unacceptably high absolute number of signals to examine.

As well as initiatives in Europe and the US, testing of methods of signal detection or data mining is occurring internationally, including, for example, in the Korean national insurance claims database, HIRA [80]. A range of methods for signal detection in healthcare databases have been examined or implemented, including measures of disproportionality as used for spontaneous reports [81], but also several other methods [82–84]. Two methods have received considerable attention, because rather than focusing on trying to generate an individual point estimate for cut-off/filtering/ranking drug–event pairs as potential emerging signals, they concentrate on visualizing how events have occurred relative to the date of a given exposure across an entire database, and therefore leverage more of the data in the record. These two methods are the chronograph, which plots IC values for each drug–outcome event stratified by the month of recording of the outcome/diagnosis before/after a given exposure, and secondly the Longitudinal Evaluation of Observational Profiles of Adverse events Related to Drugs (LEOPARD). For the chronograph, the visualization allows one to find interesting patterns at any time relative to exposure [85,86], and heuristics based on comparison of multiple potential exposed and nonexposed windows have been developed [87]. Similarly, LEOPARD is a method that can be used to attempt to automatically discard false drug–event associations caused when a drug was initiated in response to the underlying disease under investigation or misclassification of the dates of the AEs, by comparing prior event prescription rates to postevent prescription rates [88]. Similar to the chronograph, LEOPARD can generate a single test statistic, or a visualization that can be used for more qualitative information on the relationship between drug and event. Initiatives exploring EHR-based signal detection systems are intended to complement, not replace, existing drug safety surveillance systems.

Data Mining in Other Emerging Data Sources

Social Media

One of the more recent data streams that has become available for interrogation for postmarketing drug evaluation is "social media" or "digital media" – both terms used interchangeably as umbrella terms for many distinct data types [89], such as social media blogging-type sites (e.g., Facebook, Twitter), disease/product-specific discussion forums, patient engagement program data, web search logs (e.g., Google, Bing), and consumer reporting of AEs. The data on the internet are available for analysis through a variety of means. Companies and academic groups "web scrape" the internet to generate anonymized datasets for analysis; secondly, some algorithms and software solutions look to analyze data on the internet directly; and lastly social media companies that host content that users upload to the social media platforms often provide de-identified datasets for analysis, and are welcoming collaborations with external

partners to analyze the large datasets that are produced. All of these diverse data sources may be anticipated to bring different strengths and limitations in the context of signal detection.

Much research is emerging on the potential application of surveillance to social media data, but it is nascent. Examples of analyses of social media data include efforts to compare concordance to quantitative outputs from spontaneous reporting systems, such as the paper by Freifeld *et al.* [90]. They looked at the concordance between AEs suspected based on Twitter posts and US spontaneous reporting, and found a strong Spearman rank correlation. The authors concluded: "Despite the public availability of these data, their appropriate role in pharmacovigilance has not been established." They also suggested that denominator-based pharmacoepidemiologic methods needed to establish baselines and thresholds for "proto AE output" gleaned from social media. Another example is White *et al.* [91], who conducted a comparison of search log outputs and an external reference set of established drug–AE pairs. The authors reported higher accuracy in identifying the OMOP "gold-standard" reference set of known drug–event pairs, compared to that achieved using the US spontaneous reports.

Another example of analysis of social media data is Pierce *et al.* [92]. In this study, 935 246 posts were harvested from Facebook and Twitter over 5 years, and an automated classifier identified 98 252 proto-AEs. The authors focused on searching for proto-AEs that resembled 10 preselected drug–event pairs considered "positive controls" as they were recent FDA postmarketing safety signals, and 6 "negative controls" by randomly pairing each product that appeared in the set of positive controls with one event that appeared in the set of positive controls. A mere 13 posts mapped to the true positive controls across the entire set of posts A, and these were selected for causality assessment of product–event pairs, the clinical assessment revealing that the 13 posts had sufficient information to warrant further investigation for 2 of the 10 possible product–event associations: dronedarone–vasculitis and Banana Boat sunscreen–skin burns. In further analysis, for one of the two product–event pairs judged by the authors as worthy of further investigation, the first report occurred in social media prior to signal detection from the US spontaneous report database, whereas the other case occurred earlier in the spontaneous reporting system. By contrast, no product–event association posts were found among the negative controls. Overall, the literature clearly does not, at this time, articulate a case for replacing traditional spontaneous reporting with social media screening, and in fact remains inconclusive regarding the supplementary value of such activity for signal detection compared to more traditional approaches.

In social media analysis one talks about "annotation"; that is, social media observations with reliable explanatory or commentary notes, which in the case of pharmacovigilance would look to articulate explicitly the chance that a social media observation represented a suspected ADR. Sarker *et al.* in their review [93] made clear that publicly available social media (annotated with relevant medical or drug identification) data were scarce, making study/methods comparisons particularly difficult. This then limits determination of the performance characteristics of social media data analysis. Hence, the extent to which social media are able to accurately highlight emerging safety issues, while minimizing both false positive and false negative findings, is unclear. Many studies on social media have focused on data extraction rather than the novel contribution of such data. Much of the performance evaluation has concentrated on correlation of social media analysis outputs with spontaneous reports, where of course a key question is what the novel contribution of such data stream analysis might be.

Another example of ongoing research is that coming out of the EU Innovative Medicines

Initiative (IMI) "WEB-RADR" project. The project is developing a mobile app for patients and healthcare professionals to report suspected ADRs to national EU regulators, and investigating the potential for publicly available social media data for identifying drug safety issues. Although this project has seemingly made progress in showing the effectiveness of using mobile phones for reporting suspected ADRs [94], the ability so far for identifying any new AE signals more effectively than spontaneous reports and other traditional data sources seems more questionable [95]. Thus, the promise of this new data source for safety surveillance has not yet been fully realized.

Medical literature

While considerable attention has been given to mining social media records, one area possibly unique to drug safety is mining the medical literature for possible identification of ADRs. Manually searching the literature to find reports or other information on possible ADRs or adverse events following immunization (AEFI) is generally time and resource consuming. To overcome this limitation, an interesting semi-automatic approach has been described by Botsis *et al*. [96], who searched for reports of rotavirus vaccines and intussusception. They used text-mining approaches to extract medical terms for primary and secondary diagnoses, cause of death, and plain symptoms from the free text in published abstracts, and then linked them to standard MedDRA codes for further analysis. Various techniques have been developed to automate knowledge extraction for providing appropriate information [97]. The MEDLINE database from the National Library of Medicine (NLM) is a leading source of scientific information. Extracting articles related to ADRs from MEDLINE using a medical subject headings (MeSH) approach has been successfully applied in several projects [32,98,99]. Most success so far has been for signal refinement of signals detected by other methods. As an example, one study [100] aimed to automate the search of publications concerning ADRs by defining the queries used to search MEDLINE and determining the required threshold for the number of extracted publications to confirm the drug–event association in the literature. The MeSH "descriptor records" and "supplementary concept records" thesaurus was used, utilizing the subheadings "chemically induced" and "adverse effects" with the "pharmacological action" knowledge. Employing a threshold of three or more extracted publications, the automated search method had a sensitivity of 90% and a specificity of 100% [100].

Longstanding Biomedical Computing Methods Emerging in Pharmacoepidemiology

Biomedical ontologies with logical classification hierarchies have emerged and played important roles in biomedical knowledge management and data integration, as compared to controlled vocabulary resources. Specifically, a biomedical ontology is a human- and computer-interpretable set of terms and semantic relations that represent all entities in a specific biomedical domain and how these terms relate to each other. A biomedical ontology is automatically computer interpretable, since the ontology is generated using a standard computer-understandable language such as the Web Ontology Language (OWL; https://www.w3.org/OWL/) [101]. As an example, the Ontology of Adverse Events (OAE) is a recent biomedical ontology designed to represent without ambiguity various AEs observed after medical interventions, including drug administration [102]. Compared to controlled vocabulary terminologies such as the Medical Dictionary for Regulatory Activities (MedDRA) and the WHO's Adverse Reaction Terminology (WHO-ART), OAE has many advantages, such as the inclusion of textual definitions for terms and references, logical axioms, and a clearly defined and widely accepted

hierarchical structure [101,102]. Instead of defining an adverse event as shown in MedDRA and WHO-ART, OAE defines an AE as a pathologic process that occurs after a medical intervention and has an unintended outcome of a symptom, a sign, or a further pathologic process (e.g., acute infection). Such an OAE definition semantically links the AE with the medical intervention (e.g., drug administration), patient records, adverse health outcome, and temporal relation between the medical intervention and health outcome [101]. Emerging studies [103–105] suggest that OAE may provide more robust hierarchical structure definitions than MedDRA in terms of AE classification.

Natural language processing (NLP) of unstructured data is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages, and, in particular, with programming computers to fruitfully process large sections of natural-language text to extract information. It is emerging as a tool to leverage underutilized data sources that can improve pharmacovigilance, including the objective of adverse drug event detection and assessment by automatically extracting both the drugs and the potential adverse effects [106]. Research has also looked to uncover information stored in electronic medical record clinical narratives, from which summary information can be shared for research, but not the full narrative. Such an approach was, for example, used in a proof-of-concept study to provide knowledge that acute liver disease (ALD) detection could be identified earlier when an NLP-uncovered clinical narrative was combined with the structured information on ALD [107].

Chemical structure activity and wider systems biology thinking also hold potential. ADRs vary widely in mechanism, severity, and populations affected, making ADR prediction and identification important public health concerns. Systems pharmacology [108], an emerging interdisciplinary field combining network and chemical biology, provides important tools to uncover and understand ADRs and may mitigate some of the drawbacks of traditional data systems such as spontaneous reporting systems. In particular, network analysis allows researchers to integrate heterogeneous data sources and quantify the interactions between biologic and chemical entities. Recent work in this area has combined chemical, biologic, and large-scale observational health data to predict ADRs in both individual patients and global populations [108]. Drugs can act on multiple protein targets, some of which can be unrelated by traditional molecular metrics, and thousands of proteins have been implicated in adverse side effects. Some ADRs are caused by modulation of a drug's primary target, and others result from nonspecific interactions of multiple reactive metabolites. In many cases, however, ADRs are caused by unintended activity at off-targets. Recent informatics methods have been tested to systematically evaluate – on a large scale – new targets to the ADRs of those drugs for which they are the primary or well-known off-targets, creating a drug–target–ADR network [109]. Examples of attempts to include chemical structure and receptor activity relationships in pharmacovigilance have been published [110–112]. In a recently published study [11], the authors investigated the mechanisms underlying the known association of antipsychotic drugs and pneumonia. First, hypothesized mechanisms underlying antipsychotic-related pneumonia were identified through a systematic literature review; thereafter, by mining public repositories of drug targets and drug safety data these mechanisms were confirmed, and other novel antipsychotic drug targets were identified though mapping biologic pathways that could link antipsychotic drug targets and off-targets to pneumonia. In general, innovative approaches for biologic substantiation of drug–AE associations may strengthen evidence on drug safety profiles and help to tailor pharmacologic therapies to patient risk factors.

The Future

Signal detection is a critical component of pharmacoepidemiology. Historically, there has been a near complete reliance on spontaneous reporting systems for signal detection [35]. While these systems have evolved, quantitative signal detection now plays an essential role in signal detection. One wonders if we are a reaching a tipping point, where signal detection in other data streams will be an important routine complement to spontaneous reporting. If this occurs, how will spontaneous reporting evolve?

As part of postmarketing surveillance, regulatory agencies and other institutions have created and maintain large collections of suspected ADR reports. However, confounding factors such as concomitant medications, patient demographics, patient medical histories, and reasons for prescribing a drug often are uncharacterized in spontaneous reporting systems. These omissions can limit the use of quantitative signal detection methods, not to mention the lack of a denominator to allow routine accurate estimation of reporting rates [113]. Improving mechanisms to detect ADRs is a key element in strengthening postmarketing drug safety surveillance. Most signal detection relies on a single information source; methods based on jointly analyzing multiple information sources are an interesting line of research [15].

Indeed, in the future, in the same way as there will be potential consideration of new mobile data (mdata) and device data streams in pharmacoepidemiology, more work will be needed to better consider how to conduct analyses and inferential implications of analyses across multiple data streams. The dangers of getting wrong answers when pooling data across trials or databases seem to be even greater when trying to utilize data from social media with spontaneous reports. It would seem sensible to report them separately, rather than to treat them as a single source of data. When they are concordant, this strengthens the evidence for (but does not prove) a causal association. When they are discordant, this weakens the argument, and investigation of possible reasons for discordance, if evident, may provide additional insight. With these two streams (social media and spontaneous reports), concordance is likely to be greater between them in comparison to EHRs, as there is at least a suspicion of a linkage between drug and AE.

The growth in approaches to developing and implementing common data models portends an increasing availability of high-quality real-world clinical data in support of research. Building on these efforts will allow a future whereby significant portions of the world population may be able to share their data for research.

Spontaneous reporting systems and clinical observations will remain essential for postmarketing drug safety evaluation and in particular signal detection, despite the known limitations. On the other hand, vast amounts of observational data, mostly healthcare databases as part of distributed networks, offer opportunities for better pharmacovigilance surveillance in the future. Use of distributed database networks for rapid drug safety signal substantiation has been consolidating thanks to the relevant contribution of Sentinel and other worldwide initiatives. In Europe, several initiatives through multiple database networks have been conducted with the aim of exploring drug safety issues, despite the substantial efforts that are required to integrate information coming from different European healthcare databases with different underlying healthcare systems, languages, coding systems, and types of collected information (Table 27.1). Use of this source to complement SRSs for signal detection concerning events which are unlikely to be reported is promising, while optimal methods which may minimize substantially the risk of spurious signal detection have not yet been developed. SRSs currently remain the best source for looking at all possible drug–AE combinations, and we do not yet see such a universal approach being implemented effectively using EHR or similar systems. Their

Table 27.1 Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project.

Table	Fields	Aarhus	ARS	HSD	IPCI	PEDIANET	PHARMO	QRESEARCH	UNIMIB
HOSP	**Main diagnosis**	ICD10	ICD9CM				IC09CM		ICD9CM
	Secondary diagnosis	ICD10	ICD9CM				ICD9CM		ICD9CM
	Procedures	ICD10	ICD9CM				ICD9CM		ICD9CM
DEATH	**Cause of death**	ICD10	ICD9CM						
GP	**Ddiagnosis**			ICD9CM and free text	ICPC and free text NL	IC09CM and free text ITA		READ	
	Specialist			Free text ITA	Free text NL	Free text ITA			
	Lab			Free text ITA				READ	
LAB	**Classification**	NPU			WCIA		WCIA		
	Result	NPU			Numbers		Numbers		

Table information: DEATH, registry of death and causes of death; GP, information recorded by general practitioners during their clinical practice; HOSP, discharge summary from hospitalizations recorded by hospitals; LAB, information obtained from laboratories.
ARS, Tuscany Regional Database (Agenzia Regionale di Sanità); HSD, Health Search Database; ICD10, International Classification of Diseases, 10th edition; ICD9CM, International Classification of Diseases, 9th edition, with clinical modification; ICPC, International Classification of Primary Care; IPCI, Integrated Primary Information Database; ITA, Italy; Lab, Laboratory; NL, The Netherlands; NPU, nomenclature, properties and units; WCIA, Werkgroep Coördinatie Informatisering en Automatisering reference model for GP information systems; UNIMIB, Universita' di Milano Bicocca.
Source: Avillach P, Coloma PM, Gini R, *et al*. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. *J Am Med Inform Assoc* 2013; **20**(1): 184–92. Reproduced by permission of Oxford University Press.

role is going to be in the second and subsequent stages of evaluation of signals, or in looking at a range of possible signals in a limited area.

In general, we need to consider healthcare databases for providing postmarketing evidence in situations where the RCT-premarketing evidence is limited. Such is the case for high-cost, innovative medicines with fast-track approval due to a lack of alternative, orphan drugs, or drugs in pediatrics and pregnancy. For example, a healthcare database may provide a substantial contribution in the context of emerging signal detection issues related to biosimilars and biologic reference products, which are expected to be an increasingly discussed topic due to a growing number of biosimilars being marketed worldwide [114].

We live in an increasingly connected world where there is more data accumulated that may have direct or indirect impacts on pharmacoepidemiology. Informatics approaches to help us access, structure, and link data sources to enhance pharmacoepidemiology strategies will become ever more important. There is clearly more work to do in the future that might help in the next generation of new, robust pharmacoepidemiologic strategies. Near real-time clinical help systems allowing interactive healthcare patient support through trusted third parties would be valuable for enriched data collection, assuming privacy and trust concerns can be appropriately addressed. Furthermore, it will be important and interesting to increasingly examine how to leverage such data for pharmacoepidemiology and support their development. More signal substantiation using informatics across other datasets will also be important to examine. Ensuring appropriate, robust generation of evidence and its communication (including explicit communication assumptions and anticipated uncertainty) through transparent approaches will be critical as more and more evidence generation occurs increasingly rapidly. To conclude, we anticipate great advances internationally in the use of data-mining and informatics approaches to inform pharmacoepidemiology, which will act as a complement and serve only to enforce the vital role of more traditional pharmacoepidemiologic approaches in ensuring the safe and appropriate use of medicines.

References

1 Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. *Comm ACM* 1996; **39**(11): 27–34.

2 Kamrani A, Rong W, Gonzalez R. A genetic algorithm methodology for data mining and intelligent knowledge acquisition. *Comput Ind Eng* 2001; **40**(4): 361–77.

3 Bate A, Reynolds RF, Caubel P. The hope, hype and reality of Big Data for pharmacovigilance. *Ther Adv Drug Saf* 2018; **9**(1): 5–11.

4 Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? *Clin Pharmacol Ther* 2017; **102**(6): 924–33.

5 Toh S. Pharmacoepidemiology in the era of real-world evidence. *Curr Epidemiol Rep* 2017; **4**(4): 262–5.

6 Trifirò G, Sultana J, Bate A. From Big Data to smart data for pharmacovigilance: the role of healthcare databases and other emerging sources. *Drug Saf* 2017; **41**(2): 143–9.

7 Edinburgh, U. What is informatics? 2017. https://www.ed.ac.uk/informatics/about/what-is-informatics (accessed May 2019).

8 Andrews EB, Margulis AV, Tennis P, West SL. Opportunities and challenges in using epidemiologic methods to monitor drug safety in the era of large automated health databases. *Curr Epidemiol Rep* 2014; **1**(4): 194–205.

9 Aronson JK, Hauben M, Bate A. Defining "surveillance" in drug safety. *Drug Saf* 2012; **35**(5): 347–57.

10 Bauer-Mehren A, van Mullingen EM, Avillach P, *et al*. Automatic filtering and substantiation of drug safety signals. *PLoS Comput Biol* 2012; **8**(4): e1002457.

11 Sultana J, Calabro M, Garcia-Serna R, *et al*. Biological substantiation of antipsychotic-associated pneumonia: systematic literature review and computational analyses. *PLoS One* 2017; **12**(10): e0187034.

12 Edwards IR, Biriell C. Harmonisation in pharmacovigilance. *Drug Saf* 1994; **10**(2): 93–102.

13 CIOMS. *Practical aspects of signal detection in pharmacovigilance*. Report of CIOMS Working Group VIII. Geneva: CIOMS, 2010.

14 Coloma PM, Trifirò G, Patadia V, Sturkenboom M. Postmarketing safety surveillance: where does signal detection using electronic healthcare records fit into the big picture? *Drug Saf* 2013; **36**(3): 183–97.

15 Harpaz R, DuMouchel W, Schuemie M, *et al*. Toward multimodal signal detection of adverse drug reactions. *J Biomed Inform* 2017; **76**: 41–9.

16 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312**(7040): 1215–18.

17 Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 2006; **174**(5): 635–41.

18 Patadia V, Schuemie M, Coloma PM, *et al*. Can electronic health records databases complement spontaneous reporting system databases? A historical-reconstruction of the association of rofecoxib and acute myocardial infarction. *Front Pharmacol* 2018; **9**: 594.

19 Patadia VK, Schuemie MJ, Coloma P, *et al*. Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. *Int J Clin Pharm* 2015; **37**(1): 94–104.

20 Waxman HA. The lessons of Vioxx: drug safety and sales. *N Engl J Med* 2005; **352**(25): 2576–8.

21 Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004; **364**(9450): 2021–9.

22 McClellan M. Drug safety reform at the FDA: pendulum swing or systematic improvement? *N Engl J Med* 2007; **356**(17): 1700–2.

23 Gagne JJ, Wang SV, Rassen JA, Schneeweiss S. A modular, prospective, semi-automated drug safety monitoring system for use in a distributed data environment. *Pharmacoepidemiol Drug Saf* 2014; **23**(6): 619–27.

24 Edwards IR. What are the real lessons from Vioxx? *Drug Saf* 2005; **28**(8): 651–8.

25 Gee J, Naleway A, Shui I, *et al*. Monitoring the safety of quadrivalent human papillomavirus vaccine: findings from the Vaccine Safety Datalink. *Vaccine* 2011; **29**(46): 8279–84.

26 Slade BA, Leidel L, Vellozzi C, *et al*. Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. *JAMA* 2009; **302**(7): 750–7.

27 Yih WK, Greene SK, Zichittella L, *et al*. Evaluation of the risk of venous thromboembolism after quadrivalent human papillomavirus vaccination among US females. *Vaccine* 2016; **34**(1): 172–8.

28 Ioannidis JP. Exposure-wide epidemiology: revisiting Bradford Hill. *Stat Med* 2016; **35**(11): 1749–62.

29 Benjamin DJ, Berger JO, Johannesson M, *et al*. Redefine statistical significance. *Nat Hum Behav* 2017: **2**(1), 6–10.

30 Noren GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Stat Methods Med Res* 2013; **22**(1): 57–69.

31 Almenoff JS, Powell G, Schaaf R, *et al*. Online signal management: a systems-based approach that delivers new analytical capabilities and operational efficiency to the practice of pharmacovigilance. *Drug Info J* 2007; **41**(6): 779–89.

32 Amardeilh F, Bousquet C, Guillemin-Lanne S, *et al*. A knowledge management platform for documentation of case reports in pharmacovigilance. *Stud Health Technol Inform* 2009; **150**: 517–21.

33 Meyboom RH, Egberts AC, Edwards IR, Hekster YA, de Koning FH, Gribnau FW. Principles of signal detection in pharmacovigilance. *Drug Saf* 1997; **16**(6): 355–65.

34 Almenoff J, Tonning JM, Gould AL, *et al*. Perspectives on the use of data mining in pharmaco-vigilance. *Drug Saf* 2005; **28**(11): 981–1007.

35 Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf* 2009; **18**(6): 427–36.

36 Goldacre B, Drysdale H, Powell-Smith A. The COMPare trials project. 2016. http://COMPare-trials.org (accessed December 2016).

37 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by *The Lancet*. *Lancet* 2008; **372**(9634): 201.

38 Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Stat Methods Med Res* 2005; **14**(5): 515–24.

39 Pearl J. Bayesianism and causality, or, why I am only a half-Bayesian. In Corfield D, Williamson J (eds.), *Foundations of Bayesianism*. New York: Springer, 2001, pp. 19–36.

40 Ioannidi, JP. How to make more published research true. *PLoS Med* 2014; **11**(10): e1001747.

41 Gruber S, Chakravarty A, Heckbert SR, *et al*. Design and analysis choices for safety surveillance evaluations need to be tuned to the specifics of the hypothesized drug–outcome association. *Pharmacoepidemiol Drug Saf* 2016; **25**(9): 973–81.

42 Wang SV, Schneeweiss S, Berger ML, *et al*. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1018–32.

43 Langan SM, Benchimol EI, Guttmann A, *et al*. Setting the RECORD straight: developing a guideline for the REporting of studies Conducted using Observational Routinely collected Data. *Clin Epidemiol* 2013; **5**: 29–31.

44 Walker AM. Orthogonal predictions: follow-up questions for suggestive data. *Pharmacoepidemiol Drug Saf* 2010; **19**(5): 529–32.

45 Gould AL. Generating and confirming hypotheses. *Pharmacoepidemiol Drug Saf* 2010; **19**(5): 533–6.

46 Davis RL. Vaccine safety surveillance systems: critical elements and lessons learned in the development of the US Vaccine Safety Datalink's rapid cycle analysis capabilities. *Pharmaceutics* 2013; **5**(1): 168–78.

47 Behrens JT. Principles and procedures of exploratory data analysis. *Psychol Methods* 1997; **2**(2): 131.

48 Bate A, Lindquist M, Edwards IR, *et al*. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998; **54**(4): 315–21.

49 DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 1999; **53**(3): 177–90.

50 van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf* 2002; **11**(1): 3–10.

51 Waller PC, Evans SJ. A model for the future conduct of pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2003; **12**(1): 17–29.

52 Finney DJ. Systemic signalling of adverse reactions to drugs. *Methods Inf Med* 1974; **13**(1): 1–10.

53 Lindquist M, Stahl M, Bate A, Edwards IR, Meyboom RH. A retrospective evaluation of a data mining approach to aid finding new adverse

drug reaction signals in the WHO international database. *Drug Saf* 2000; **23**(6): 533–42.

54 Roux E, Thiessard F, Fourrier A, Begaud B, Tubert-Bitter P. Evaluation of statistical association measures for the automatic signal generation in pharmacovigilance. *IEEE Trans Inf Technol Biomed* 2005; **9**(4): 518–27.

55 DuMouchel W, Pregibon D. Empirical Bayes screening for multi-item associations. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2001, pp. 67–76.

56 Caster O, Noren GN, Madigan D, Bate A. Logistic regression in signal detection: another piece added to the puzzle. *Clin Pharmacol Ther* 2013; **94**(3): 312.

57 Hopstadius J, Noren GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. *Drug Saf* 2008; **31**(11): 1035–48.

58 Meyboom RH, Lindquist M, Egberts AC, Edwards IR. Signal selection and follow-up in pharmacovigilance. *Drug Saf* 2002; **25**(6): 459–65.

59 Stricker BH, Tijssen JG. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol* 1992; **45**(10): 1177–84.

60 Moore N, Kreft-Jais C, Haramburu F, *et al*. Reports of hypoglycaemia associated with the use of ACE inhibitors and other drugs: a case/non-case study in the French pharmacovigilance system database. *Br J Clin Pharmacol* 1997; **44**(5): 513–18.

61 Caster O, Juhlin K, Watson S, Noren GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. *Drug Saf* 2014; **37**(8): 617–28.

62 Chan CL, Rudrappa S, Ang PS, Li SC, Evans SJW. Detecting signals of disproportionate reporting from Singapore's spontaneous adverse event reporting system: an application of the sequential probability ratio test. *Drug Saf* 2017; **40**(8): 703–13.

63 Chandler RE, Juhlin K, Fransson J, Caster O, Edwards IR, Noren GN. Current safety concerns with human papillomavirus vaccine: a cluster analysis of reports in VigiBase(R). *Drug Saf* 2017; **40**(1): 81–90.

64 Crooks CJ, Prieto-Merino D, Evans SJ. Identifying adverse events of vaccines using a Bayesian method of medically guided information sharing. *Drug Saf* 2012; **35**(1): 61–78.

65 de Bie S, Verhamme KM, Straus SM, Stricker BH, Sturkenboom MC. Vaccine-based subgroup analysis in VigiBase: effect on sensitivity in paediatric signal detection. *Drug Saf* 2012; **35**(4): 335–46.

66 Gould AL. Accounting for multiplicity in the evaluation of "signals" obtained by data mining from spontaneous report adverse event databases. *Biom J* 2007; **49**(1): 151–65.

67 Heeley E, Waller P, Moseley J. Testing and implementing signal impact analysis in a regulatory setting: results of a pilot study. *Drug Saf* 2005; **28**(10): 901–6.

68 Noren GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. *Data Min Knowl Discov* 2007; **14**(3): 305–28.

69 Noren GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug–drug interaction surveillance. *Stat Med* 2008; **27**(16): 3057–70.

70 Orre R, Bate A, Lindquist M. Bayesian neural networks used to find adverse drug combinations and drug related syndromes. In Malmgren H, Borga M, Niklasson L (eds.), *Artificial Neural Networks in Medicine and Biology*. New York: Springer, 2000, pp. 215–20.

71 Orre R, Bate A, Noren GN, Swahn E, Arnborg S, Edwards IR. A Bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets. *Int J Neural Syst* 2005; **15**(3): 207–22.

72 Pariente A, Avillach P, Salvo F, *et al*. Effect of competition bias in safety signal generation: analysis of a research database of spontaneous

reports in France. *Drug Saf* 2012; **35**(10): 855–64.

73 Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med* 2012; **31**(30): 4401–15.

74 Pacurariu AC, Straus SM, Trifirò G, *et al*. Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. *Drug Saf* 2015; **38**(12): 1201–10.

75 Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**(7125): 140–4.

76 Vandenbroucke JP. The HRT controversy: observational studies and RCTs fall in line. *Lancet* 2009; **373**(9671): 1233–5.

77 Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004; **363**(9422): 1728–31.

78 Lievre M, Cucherat M, Leizorovicz A. Pooling, meta-analysis, and the evaluation of drug safety. *Curr Control Trials Cardiovasc Med* 2002; **3**(1): 6.

79 Xu Y, Zhou X, Suehs BT, *et al*. A comparative assessment of Observational Medical Outcomes Partnership and Mini-Sentinel common data models and analytics: implications for active drug safety surveillance. *Drug Saf* 2015; **38**(8): 749–65.

80 Choi NK, Chang Y, Choi YK, Hahn S, Park BJ. Signal detection of rosuvastatin compared to other statins: data-mining study using national health insurance claims database. *Pharmacoepidemiol Drug Saf* 2010; **19**(3): 238–46.

81 Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res* 2013; **22**(1): 39–56.

82 Brown JS, Petronis KR, Bate A, *et al*. Drug adverse event detection in health plan data using the gamma Poisson shrinker and comparison to the tree-based scan statistic. *Pharmaceutics* 2013; **5**(1): 179–200.

83 Cook AJ, Tiwari RC, Wellman RD, *et al*. Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the Mini-Sentinel pilot. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 72–81.

84 Lai EC, Pratt N, Hsieh CY, *et al*. Sequence symmetry analysis in pharmacovigilance and pharmacoepidemiologic studies. *Eur J Epidemiol* 2017; **32**(7): 567–82.

85 Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov* 2010; **20**(3): 361–87.

86 Cederholm S, Hill G, Asiimwe A, *et al*. Structured assessment for prospective identification of safety signals in electronic medical records: evaluation in the health improvement network. *Drug Saf* 2015; **38**(1): 87–100.

87 Noren GN, Bergvall T, Ryan PB, Juhlin K, Schuemie MJ, Madigan D. Empirical performance of the calibrated self-controlled cohort analysis within temporal pattern discovery: lessons for developing a risk identification and analysis system. *Drug Saf* 2013; **36**(Suppl 1): S107–S121.

88 Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. *Pharmacoepidemiol Drug Saf* 2011; **20**(3): 292–9.

89 Bate A, Reynolds RF, Caubel P. *The Hope, Hype and Reality of Big Data for Pharmacovigilance*. London: Sage, 2017.

90 Freifeld CC, Brownstein JS, Menone CM, *et al*. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf* 2014; **37**(5): 343–50.

91 White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated

data on the internet. *Clin Pharmacol Ther* 2014; **96**(2): 239–46.

92 Pierce CE, Bouri K, Pamer C, *et al*. Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts. *Drug Saf* 2017; **40**(4): 317–31.

93 Sarker A, Ginn R, Nikfarjam A, *et al*. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform* 2015; **54**: 202–12.

94 Huesch MD. Commercial online social network data and statin side-effect surveillance: a pilot observational study of aggregate mentions on Facebook. *Drug Saf* 2017; **40**(12): 1199–204.

95 Noren N. Needles in hay: use of machine learning to identify references to drugs and adverse events in social media. *Drug Saf* 2017; **40**(10): 1037–8.

96 Botsis T, Foster M, Kreimeyer K, Pandey A, Forshee R. Monitoring biomedical literature for post-market safety purposes by analyzing networks of text-based coded information. *AMIA Jt Summits Transl Sci Proc* 2017; 66–75.

97 Mendonca EA, Cimino JJ. Automated knowledge extraction from MEDLINE citations. *Proc AMIA Symp* 2000; 575–9.

98 Avillach P, Joubert M, Thiessard F, *et al*. Design and evaluation of a semantic approach for the homogeneous identification of events in eight patient databases: a contribution to the European EU-ADR project. *Stud Health Technol Inform* 2010; **160**(Pt 2): 1085–9.

99 Garcelon N, Mougin F, Bousquet C, Burgun A. Evidence in pharmacovigilance: extracting adverse drug reactions articles from MEDLINE to link them to case databases. *Stud Health Technol Inform* 2006; **124**: 528–33.

100 Avillach P, Dufour JC, Diallo G, *et al*. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. *J Am Med Inform Assoc* 2013; **20**(3): 446–52.

101 Wang L, Li M, Xie J, Cao Y, Liu H, He Y. Ontology-based systematical representation and drug class effect analysis of package insert-reported adverse events associated with cardiovascular drugs used in China. *Sci Rep* 2017; **7**(1): 13819.

102 He Y, Sarntivijai S, Lin Y, *et al*. OAE: the ontology of adverse events. *J Biomed Semantics* 2014; **5**: 29.

103 He Y. Ontology-based vaccine and drug adverse event representation and theory-guided systematic causal network analysis toward integrative pharmacovigilance research. *Curr Pharmacol Rep* 2016; **2**(3): 113–28.

104 Sarntivijai S, Xiang Z, Shedden KA, *et al*. Ontology-based combinatorial comparative analysis of adverse events associated with killed and live influenza vaccines. *PLoS One* 2012; **7**(11): e49941.

105 Xie J, Codd C, Mo K, He Y. Differential adverse event profiles associated with BCG as a preventive tuberculosis vaccine or therapeutic bladder cancer vaccine identified by comparative ontology-based VAERS and literature meta-analysis. *PLoS One* 2016; **11**(10): e0164792.

106 Harpaz R, Callahan A, Tamang S, *et al*. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf* 2014; **37**(10): 777–90.

107 Walker AM, Zhou X, Ananthakrishnan AN, *et al*. Computer-assisted expert case definition in electronic health records. *Int J Med Inform* 2016; **86**: 62–70.

108 Boland MR, Jacunski A, Lorberbaum T, Romano JD, Moskovitch R, Tatonetti NP. Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms. *Wiley Interdiscip Rev Syst Biol Med* 2016; **8**(2): 104–22.

109 Lounkine E, Keiser MJ, Whitebread S, *et al*. Large-scale prediction and testing of drug

activity on side-effect targets. *Nature* 2012; **486**(7403): 361–7.

110 Alvager T, Smith TJ, Vijai F. Neural-network applications for analysis of adverse drug reactions. *Biomed Instrum Technol* 1993; **27**(5): 408–11.

111 Vanderwall DE, Yuen N, Al-Ansari M, *et al.* Molecular clinical safety intelligence: a system for bridging clinically focused safety knowledge to early-stage drug discovery – the GSK experience. *Drug Discov Today* 2011; **16**(15–16): 646–53.

112 Low YS, Caster O, Bergvall T, *et al.* Cheminformatics-aided pharmacovigilance: application to Stevens–Johnson syndrome. *J Am Med Inform Assoc* 2016; **23**(5): 968–78.

113 Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012; **4**(125): 125–31.

114 Kurki P, van Aerts L, Wolff-Holz E, Giezen T, Skibeli V, Weise M. Interchangeability of biosimilars: a European perspective. *BioDrugs* 2017; **31**(2): 83–91.

28

Pharmacoepidemiologic Research on Drugs of Abuse

Jana McAninch, Alex Secora, Cynthia Kornegay, and Judy Staffa

Division of Epidemiology, Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

Although pharmaceutical drug abuse is not a new phenomenon, there has been an explosion of interest and research in this area in recent years, largely due to the unprecedented crisis of opioid abuse, addiction, and overdose in the US [1], and to a lesser extent in other developed countries. Beginning in the 1990s, a dramatic rise in prescription opioid overdoses paralleled increases in the prescribing of these drugs in the US, particularly for chronic noncancer pain [2]. The urgency around the opioid epidemic has generated a need to identify and understand emerging trends, identify populations and geographic areas at greatest risk, develop evidence-based interventions, and evaluate the impact of those interventions. There is a need to understand not only the scope and patterns of the problem at the population level, but also individual-level risk factors and the role of pharmaceutical products in substance use disorders and drug overdoses.

From a regulatory standpoint, there is a need to understand misuse, abuse, and related risks associated with specific pharmaceutical products. Related adverse clinical outcomes – including addiction, overdose, and death – may involve not only the pharmaceutical drug of interest, but also other drugs, including illicit street drugs and counterfeit drugs. The US Food and Drug Administration (FDA) has identified five outcomes – misuse, abuse, addiction, overdose, and death – as issues necessitating both postmarketing safety studies and risk evaluation and mitigation strategies (REMS) for opioid analgesics. Although terminology and outcome definitions vary, the FDA has used the following definitions for "abuse" and "misuse":

> Abuse is defined as the intentional, nontherapeutic use of a drug product or substance, even once, to achieve a desirable psychological or physiological effect. Abuse is not the same as misuse, which refers to the intentional therapeutic use of a drug product in an inappropriate way and specifically excludes the definition of abuse. [3]

In some other contexts, for example in the National Survey on Drug Use and Health [4], the terms "misuse" and "nonmedical use" are used more broadly to refer to any use of a drug other than as directed by a healthcare provider. Terminology around drug addiction also varies, and has evolved, particularly with the release of the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-V). The DSM-V

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

contains revised diagnostic categories and criteria for substance use disorders [5]. Currently approved opioid analgesic labeling describes "addiction" as:

> A cluster of behavioral, cognitive, and physiological phenomena that develop after repeated substance use and include: a strong desire to take the drug, difficulties in controlling its use, persisting in its use despite harmful consequences, a higher priority given to drug use than to other activities and obligations, increased tolerance, and sometimes a physical withdrawal … Abuse and addiction are separate and distinct from physical dependence and tolerance. [6]

Pharmaceutical drug overdose may be broadly defined as the accidental or intentional use of a drug in an amount higher than is indicated; however, in practice, overdose is often used synonymously with the constellation of adverse effects associated with supratherapeutic use for a particular drug class. These include, for example, severe respiratory and central nervous system (CNS) depression, and, in the case of opioid overdose, possibly death. For simplicity, throughout this chapter we will sometimes use the general term "abuse" to refer to abuse and these related outcomes.

Although many fundamental principles of pharmacoepidemiology apply to the study of pharmaceutical drug abuse, there are also unique challenges and approaches, which are outlined in this chapter. Some of these challenges arise due to the diversion of drugs with abuse potential from the intended patient to others, making complete exposure difficult to ascertain; these individuals are then also at risk for adverse outcomes (Figure 28.1). Outcomes in this area can be difficult to measure, as they often relate to covert behaviors, lack clear definitions, and are not fully captured in traditional healthcare data systems. Therefore, a "mosaic" approach is frequently employed, where multiple studies using a variety of data sources and methods are qualitatively synthesized to answer an abuse-related research or regulatory question.

Finally, the landscape of controlled substance prescribing and substance abuse is ever-changing,

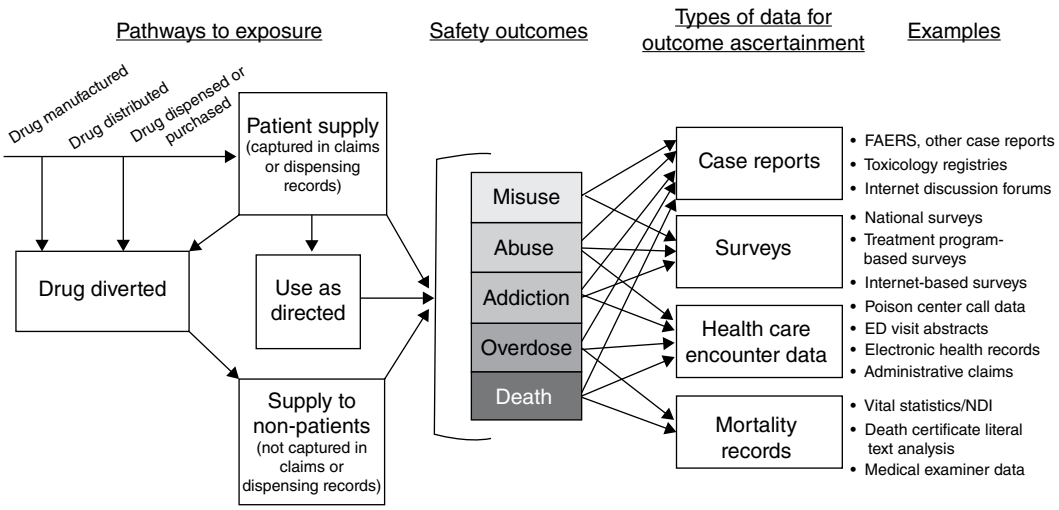


Figure 28.1 Ascertaining exposure and outcomes in studies of pharmaceutical drug abuse. ED, emergency department; FAERS, Food and Drug Administration Adverse Event Reporting System; NDI, National Death Index.

with myriad interventions being implemented at federal, state, and local levels, and with pharmaceutical and illicit drug abuse patterns intersecting in complicated ways that are often difficult to fully characterize. Opioids represent the focus of much of the current research in this rapidly evolving field and are the focus of most of this chapter. Many of the methodologic questions around studying opioid abuse apply to other drugs with abuse potential as well. Prescription stimulants, anxiolytics, sedative/hypnotics, and other drugs with CNS activity have safety concerns related to misuse, abuse, and addiction.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

The types of clinical problems that need to be addressed by pharmacoepidemiologic research in this area can be grouped into three broad categories:

- *Signal detection*: to determine whether a drug is being abused in the community, and to understand the associated clinical consequences.
- *Descriptive quantitative assessment*: to determine the scope, demographic and geographic patterns, and trends in use, abuse, and related morbidity and mortality for a drug, class, or drug combination, or to quantify the risk of abuse and related outcomes associated with a drug or class, among those dispensed the drug(s).
- *Analytic studies*: to determine the relative risk of abuse and related outcomes using analytic epidemiologic study designs with comparator groups; for example, to evaluate risk factors for abuse and related adverse outcomes, or the effectiveness of an intervention designed to reduce abuse.

One pharmacologic concept that may take on different connotations when studying drugs of abuse is that of type A and type B reactions. Type A reactions refer to adverse reactions that are the result of an exaggerated but otherwise usual pharmacologic effect of the drug. Type A reactions tend to be common, dose related, predictable, and historically considered to be less serious. Type B reactions refer to aberrant effects, which tend to be uncommon, not related to dose, unpredictable, and potentially more serious. For example, opioid-associated respiratory depression could be categorized as a type A reaction, as it is a predictable dose-related effect that can be managed by reducing the opioid dose. However, in contrast to many type A reactions with other drugs, opioid-associated respiratory depression is frequently life threatening or fatal, and often occurs in association with abuse. The more idiosyncratic type B reactions can also arise in cases of abuse, for example due to the intravenous injection of excipients contained in a product formulation indicated only for oral use. Whereas detecting previously unknown type B reactions represents the major focus of many pharmacoepidemiologic studies, type A reactions – such as opioid overdose – are a primary public health concern and a target of research for drugs with abuse potential.

Signal Detection

The abuse potential of a new drug product with CNS activity is evaluated prior to approval and market introduction based on established criteria that determine scheduling per the Controlled Substance Act (CSA) [7,8]. However, premarket assessments cannot always accurately predict abuse in the community, where sociocultural and economic forces become manifest. Therefore, epidemiologic data are sometimes used to examine whether a drug is being abused in the postapproval setting, and to identify associated adverse clinical consequences. Abuse signals can also help identify potential risk factors or high-risk populations.

Quantitative Descriptive Assessment

As in other areas of drug safety evaluation, quantitative assessment of an established signal may be needed. Such an assessment may seek to quantify the risk of abuse and related outcomes among those exposed to a drug, to better understand the progression from therapeutic use to addiction, or to conduct ongoing surveillance to understand the scope, patterns, and trends in abuse of a drug product in the population. Measuring and tracking drug-involved mortality form one of the central responsibilities of public health, particularly amid the national opioid crisis; however, the complexity of attributing deaths to specific drugs and the variability in death certificate documentation present significant challenges [9].

Some of the assumptions of classic pharmacoepidemiologic studies may not be met when studying abuse and related adverse effects of pharmaceutical products, and modifications to study approaches and designs are often needed. For example, one of the basic principles of studying drug safety issues associated with prescription drugs is that patients use the drugs prescribed to them, and that the date of a dispensed medication approximates the start of the exposure period. However, in the case of drugs with abuse potential, the individual prescribed the drug may give or sell it to others, and may use additional drugs obtained from sources other than their own prescribed medication (i.e., diversion). Therefore, adverse outcomes associated with exposure to a drug can occur in individuals other than those to whom the drugs were prescribed and dispensed. In addition, because drug abuse is generally a covert activity, abuse-related outcomes may not come to medical attention, and can be difficult to measure using traditional pharmacoepidemiologic sources such as administrative insurance claims and electronic medical record data. Many serious social, legal, psychiatric, and physical adverse effects can also occur and yet not be documented in medical records or claims.

Analytic Studies

Finally, there are questions that require analytic study designs, including comparator groups and formal hypothesis testing. Clinical and policy questions often arise regarding risk factors for adverse outcomes such as addiction and overdose. For example, risk factors may include patient characteristics or co-morbidities, drug dose, or concomitant use of other drugs. Another common study question is whether specific interventions – for example, mandatory use of state prescription drug monitoring programs (PDMPs), prescribing guidelines, REMS (see Chapter 24), and opioid formulations with properties designed to deter abuse – are effective in mitigating abuse and related adverse outcomes. To answer these questions, studies may formally compare rates of abuse and related outcomes across drug products or classes, geographic subgroups, or time periods. A wide variety of data – including data not routinely used in pharmacoepidemiology – are used for these types of investigations; however, all have substantial limitations and challenges, as discussed in the next section. The design and interpretation of analytic studies in this area require a thorough understanding of the strengths and limitations of the underlying data and methods, particularly when attempting to make causal inferences.

Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

Despite some overlap, the areas of inquiry described – signal detection, descriptive quantitative assessment, and epidemiologic studies involving comparator groups and formal hypothesis testing – generally require different types of data and methods. As the study questions become more quantitative and analytic, investigations require more systematic and rigorous data collection and statistical methods.

Again, many of the methodologic problems encountered in this subject area are not unique, but some challenges are relatively specific to the study of pharmaceutical drug abuse and related safety outcomes. Both drug abuse patterns and the data systems available to study them vary widely across countries. This chapter will primarily discuss US examples; however, the types of data and methods covered may be relevant for studies in other countries as well.

Data Sources

Signal Detection

Several data sources may be of value for identifying and characterizing postmarket abuse signals for pharmaceutical products. The FDA's Adverse Event Reporting System (FAERS, see Chapter 10) regularly receives reports describing abuse of approved drug products, intentional use via unintended routes of administration, manipulation of the product for the purposes of abuse, and adverse effects associated with these behaviors. Published case reports, case series, and media reports may also play a role in abuse signal detection for marketed drugs. Spontaneous reporting data in this area share many limitations common to evaluation of other safety signals. For example, reports do not always contain enough detail to properly evaluate an event as being abuse related, and many factors can influence whether an event will be reported. Because misuse and abuse are typically covert behaviors, adverse events associated with these behaviors may be even less likely to be reported or brought to medical attention, exacerbating underascertainment of these events in spontaneous reporting systems. Medical toxicology case registries, such as the Toxicology Investigators Consortium (ToxIC) [10], may provide more detailed clinical information and have value for abuse-related signal detection as well.

An emerging source of abuse signal detection is internet-based recreational drug use message boards, discussion forums, social media, and other web-based portals, where individuals discuss their experiences with abusing various pharmaceutical products and other drugs, often comparing preferred routes of abuse and methods for product tampering to increase availability of the active pharmaceutical ingredient. Several proprietary web monitoring programs collect and analyze these types of data [11,12]. In another online data source, individuals anonymously report prices they paid or heard were paid for various drugs [13]. Reports from law enforcement that a drug is being diverted or sold on the street may also serve as a signal that a drug is being abused [14]. Anecdotal reports from sentinel cohorts of known substance abusers (e.g., those using syringe exchange services) may also be useful in elucidating emerging trends in pharmaceutical abuse. As with traditional spontaneous reporting data, these data generally do not support quantitative estimation of abuse or comparison of abuse liability across products or time periods.

Quantitative Descriptive Assessment and Analytic Studies

Quantitative assessments of abuse-related risk, and the scope, pattern, and trends in abuse-related adverse outcomes, require more robust data collection systems, ideally using validated measures and probability sampling methods. Hypothesis testing studies use many of the same data sources, but require even greater attention to data limitations, potential bias, and measured and unmeasured confounding to facilitate causal inferences. Some of the data described in the previous section may provide contextual or supporting information for these types of investigations.

Nationally representative surveys, for example the National Survey on Drug Use and Health (NSDUH) [4] and Monitoring the Future (MTF) [15], can play an important role in both descriptive and analytic (i.e., hypothesis-testing) studies. NSDUH captures information on self-reported

drug-taking behaviors, including the nonmedical use of pharmaceuticals, in the general US population, while MTF focuses on drug-taking behaviors in secondary school students and young adults. National surveys are valuable for studying broad trends in the misuse of pharmaceutical products with known abuse potential; however, the lack of product- and formulation-specific information presents a substantial limitation in the utility of these data for estimating the misuse and abuse of specific drug products. Smaller surveys may target specific populations and drug classes – for example, surveys of college students to study misuse and abuse of stimulant drugs used to treat attention deficit disorder [16] – but the generalizability of these findings is limited. An emerging approach to substance abuse research is the use of "opt-in" internet survey panels, where a survey administration company recruits individuals to join an online panel from which participants may opt to complete various surveys. Internet panel surveys have cost advantages over traditional in-person, school-based, and telephone-based surveys, and they can be more easily tailored to examine specific products or questions [17,18]. However, these surveys do not use probability sampling techniques, and further work is needed to understand the underlying population represented, the selection forces operating, and how these might be changing over time given the rapid evolution of internet and social media use. Another method is to recruit enriched populations through web-based drug abuse discussion forums, by inviting visitors to these websites to complete online surveys exploring specific drug abuse-related questions. The relative ease of modifying these internet-based questionnaires is an obvious advantage, but rigorous questionnaire development and validation are still needed to ensure high-quality data.

Several existing surveillance systems collect information from individuals being assessed for, or who are entering, substance use disorder treatment. One of these is the Treatment Episode Data Set (TEDS). TEDS is an admission-based system that collects and aggregates information on demographic and substance use disorder characteristics among those admitted to substance use disorder treatment facilities that report to state administrative data systems [19]. Similar to national surveys of the general population, however, product- and formulation-specific information is generally not available. Several proprietary surveillance programs in the US collect more detailed data from people entering or being assessed for treatment, including self-reported recent abuse of specific pharmaceutical products and illicit drugs [20,21]. The primary strengths of these proprietary surveillance programs are their comprehensiveness with respect to specific products and behaviors, including route of abuse. The enriched nature of these study populations also improves the precision of abuse estimates for products with lower utilization. However, all of these data systems are limited, in that participant selection and outcome ascertainment depend on an individual seeking or being referred to treatment, and numerous factors (e.g., judicial referral policies and availability of and funding for treatment) affect the probability that an individual who is abusing is included in the sample. Furthermore, the nonrepresentative and dynamic nature of these samples can bias between-drug comparisons, and changes in the distribution of participating assessment and treatment sites over time can bias time-dependent analyses, such as pre–post comparisons and trends, as sites drop in and out of the surveillance networks [22,23]. Finally, all surveys are subject to the misclassification inherent to self-reported data, particularly with respect to respondents' ability to correctly identify specific products and formulations that they have abused. The extent of such misclassification and the degree to which it may be differential – for example, depending on how long a product has been on the market or the manner or order in which it is presented in a survey – are largely unknown and represent important areas for further research.

Call data from poison control centers (PCCs) and data extracted from emergency department (ED) visit records can play a role in understanding the scope, trends, and patterns of pharmaceutical drug abuse and related adverse outcomes, and these data are often used in comparative analytic evaluations as well. PCC data include reports of drug and other substance exposures and consequent adverse effects, based on calls from exposed individuals or someone caring for them, including health professionals [24]. PCC data have several strengths for studying the abuse of pharmaceutical products, including the capture of meaningful abuse-related outcomes, recording of drug exposure at the most detailed level possible based on the information provided, national or near-national geographic coverage, and ability to analyze in near real time. An important limitation of these data is that an unknown proportion of abuse-related events result in a call to a PCC, and it is unclear which factors might influence whether a call is made. Furthermore, these factors might vary over time or across drugs, making interpretation of direct drug–drug comparisons and time trends difficult. In addition, the ability to distinguish specific product brands and formulations (e.g., immediate versus extended release) is limited. Finally, out-of-hospital deaths are unlikely to be captured in PCC data, a critical limitation for potent opioids and other drugs with the potential for rapidly fatal overdose. Some research suggests that trends in the number of opioid-related deaths in PCC data correlate with trends in national mortality data [25,26], but more work is needed in this area to better understand the fraction of poisoning and overdose deaths likely to be captured in PCC data, and whether this varies across drugs or drug classes, with other case characteristics, or over calendar time.

As with PCC call data, data captured in ED visit records provide information not only about abuse behaviors, but also about morbidity and burden to the healthcare system. Again, the completeness and accuracy of these data depend on the completeness and accuracy of the medical records themselves. Currently available national ED surveillance systems are limited in the US. Data collection ended in 2011 for the Drug Abuse Warning Network (DAWN), in which data on drug-involved ED visits were manually abstracted from a probability sample of hospitals [27]. The National Electronic Injury Surveillance System: Cooperative Adverse Drug Events Surveillance System (NEISS-CADES) surveillance network, operated by the Centers for Disease Control and Prevention in collaboration with the Consumer Product Safety Commission, uses similar methods to capture ED cases and generate national estimates for ED visits related to adverse drug events. In 2016, NEISS-CADES began including cases related to drug abuse, withdrawal, or attempted self-harm. The Nationwide Emergency Department Sample (NEDS) is a surveillance system managed by the Agency for Healthcare Research and Quality (AHRQ) [28]. NEDS can be used to generate national estimates of hospital-based ED visits based on ICD-9-CM and ICD-10-CM diagnosis and external cause-of-injury codes; however, information on the involvement of specific drug products is generally not available, and drug molecule/class involvement information is limited by the International Classification of Diseases (ICD) coding system.

Electronic health records and administrative claims data have been used to assess the prevalence of misuse and abuse – or aberrant behaviors suggestive of misuse or abuse – among patients using opioids [29]. Various efforts have been made to create algorithms identifying opioid misuse and abuse using electronic healthcare data; however, both conceptual and operational outcome definitions vary widely [30], and the lack of a true gold standard complicates the evaluation of algorithm performance [31]. Administrative insurance claims may include diagnosis codes and treatment claims for substance use disorder (SUD);

however, identifying these outcomes in claims requires (i) a patient's interaction with the medical system; (ii) a practitioner's recognition and documentation; and (iii) submission of a claim for insurance reimbursement. One or more of these requirements is likely lacking in a high proportion of substance use disorder cases. Drug overdose may be ascertained in claims if medical treatment is sought; however, algorithms need to be sufficiently validated in the data source being used, particularly if they are being used in formal hypothesis-testing studies. Also, due to the high potential for out-of-hospital death, data must include or be linked to a reliable source of mortality data, such as the National Death Index (NDI) [32]. Despite their limitations, if linked to death data claims data may be useful in evaluating patient- and drug-related risk factors for overdose, for example demographic characteristics, co-morbidities, and concomitant use of other drugs. However, socioeconomic, psychological, and legal factors may play particularly strong roles as confounders of associations between drug exposure and abuse-related adverse outcomes, factors not typically captured well in claims data. Claims-based studies are further limited in this area, since pharmaceutical drug abuse and associated adverse outcomes often occur in those who obtain the drugs from a source (e.g., friend, family member, dealer) other than their own prescription, and, as is the case with other pharmaceuticals, self-paid dispensings and over-the-counter (OTC) products are generally not captured.

Vital statistics systems provide information on drug-involved mortality, including deaths that involve pharmaceutical products. These are primarily employed in descriptive analyses, but are sometimes used to evaluate interventions [33] or compare overdose risk across drugs [34]. In the US, the National Vital Statistics System (NVSS) Multiple Cause of Death data files contain national mortality and population data based on death certificates for US residents. Each death certificate contains a single underlying cause of death, up to 20 multiple causes, and demographic data. These data have several important limitations. First, involved drugs are grouped into broader categories based on the ICD-10 Injury Mortality Diagnosis Matrix [35]. Second, documentation of specific drugs on death certificates is entirely dependent on the certifier (e.g., medical examiner or coroner), and death investigation practices and documentation of involved drugs vary widely based on resources, training, location, workload, and other factors. A newly available mortality data resource is the Drug-Induced Mortality database, which links NVSS files to literal text information from death certificates to identify mentions of specific drugs [36], allowing for a more granular analysis of specific drugs involved in deaths, if they are noted [37]. Finally, medical examiner and coroner records, in comparison with death certificates, may be more timely and contain more detail on drug overdose deaths. These data are sometimes made available to researchers through agreements with individual state health departments, and efforts are underway to make these data more available nationally [38].

Data gathered from law enforcement and other agencies on drug diversion cases can be used to identify pharmaceutical products found outside of controlled distribution channels [14], and databases containing information on chemical analyses of drugs seized by law enforcement are another source for monitoring drug trafficking [39]. Diversion information is important for monitoring emerging trends to guide policy, public health, and enforcement activities. It can also provide context and aid in interpreting results of other studies in this area. However, drug diversion data are not direct measures of abuse or related clinical outcomes, but rather measures of law enforcement activity. It is unclear how funding or local law enforcement priorities may influence the number of identified drug diversion cases, the drugs on which investigators focus their efforts, and how these factors may vary over time.

Drug utilization data – from state PDMPs and other retail dispensing and administrative claims databases – are widely available and often analyzed to understand controlled substance prescribing and use patterns, including use in different age groups and geographic regions, changes over time, and concomitant use of multiple classes of controlled substances (e.g., opioids and benzodiazepines). These data are sometimes employed, either with or without linkage to additional data, to try to identify misuse, abuse, or diversion by identifying dispensing patterns suggestive of "doctor/pharmacy shopping," discussed further in the analytic methods section. While these analyses may identify some patients at elevated risk of misuse, abuse, addiction, and/or diversion, their ability to discriminate therapeutic from nontherapeutic use is a subject of ongoing research. Therefore, while "doctor/pharmacy shopping" data may have some value, they should not be interpreted as a proxy outcome measure for abuse.

Analytic Methods

Signal Detection

Methods for abuse signal detection are similar to those used to examine other safety signals, with the case series being the predominant study design. Quantitative approaches, such as Bayesian data-mining algorithms and disproportionate reporting rate analyses (see Chapter 27), can also be used for abuse-related adverse events in spontaneous adverse event reporting systems. These types of analyses have recently been conducted for gabapentinoid abuse in European data systems [40–42]. Calculating more traditional reporting rates for an abuse-related adverse event – using a population-based utilization estimate as the denominator – is also possible, but relies heavily on untestable assumptions.

Web monitoring programs typically employ a mixed quantitative and qualitative approach to studying the abuse of specific products, for example categorizing and quantifying discussion forum postings as positive, negative, or neutral experiences, or identifying posts that discuss recipes for tampering with a product for abuse purposes [43,44]. These types of analyses can be valuable for signal detection and generating hypotheses for formal studies, but they have limited utility for quantifying abuse or making comparisons across drugs or time periods. The use of websites is constantly evolving and the quality of information is highly variable.

Quantitative Descriptive Assessment

Quantitative assessment requires careful consideration of sample representativeness and appropriate denominators. Data from PCCs, ED records, surveys, vital statistics mortality databases, and other types of population data are generally normalized using one or more different denominators, depending on the study question. The first type of denominator is derived from the total study population (e.g., the number surveyed within a given time period) or the population of the study coverage area (e.g., a population covered by participating PCCs or treatment program sites). Estimates using population-based denominators are useful for understanding the scope and public health burden of abuse for different drugs in a population. The large, nationally representative surveys use survey weights to generate national prevalence estimates. Therefore, repeat cross-sectional estimates can be employed to conduct ecologic time-series assessments, often employing trend analysis methods such as time-series regression analysis or Joinpoint regression [45]. Surveys using convenience samples generally cannot be utilized to generate national abuse prevalence estimates, and they present challenges in evaluating trends over time due to nonrandom changes in the study population distribution.

Like other safety outcomes, the frequency of abuse and related outcomes depends in part on the extent to which a drug is prescribed.

Therefore, the second type of denominator incorporates some measure of drug utilization to account for dissimilar or changing levels of prescribed availability. The measure of prescribed availability used (e.g., unique recipients of drug, prescriptions dispensed, tablets dispensed, total morphine milligram equivalents dispensed) can dramatically affect the relative magnitude of the abuse estimates for different drugs [46,47]. Again, the most appropriate metric depends on the question being asked. For example, when comparing abuse rates for drugs with widely varying numbers of tablets per prescription (e.g., immediate-release opioids), the total number of tablets dispensed may be the best denominator, as each tablet can be viewed as an opportunity for diversion and abuse. One might also wish to examine differential abuse rates of different opioid active moieties (e.g., hydrocodone, oxycodone, tramadol), accounting for the different analgesic potencies and dosage strengths available. In this case, the most useful denominator might be total morphine milligram equivalents dispensed. As with population denominators, the appropriate coverage area for utilization denominators can be difficult to define in studies using nonrepresentative convenience samples.

Finally, using some subset of the study population as the denominator can also generate meaningful measures of abuse behaviors. For example, the number of individuals surveyed who report *any* use of a drug (therapeutic or nontherapeutic) may serve as a denominator, allowing estimation of the proportion of those who have *used* a particular drug who also have *abused* it in a given time period. Or, one might estimate the proportion of individuals abusing a particular drug who report doing so through specific routes such as chewing, snorting, or injecting. This type of analysis is sometimes referred to as a drug's "route-of-abuse profile."

Cohort studies can provide valuable descriptive data as well, for example linking electronic healthcare data to death certificate data in a retrospective cohort design to study rates of fatal and nonfatal overdose among patients receiving chronic opioid therapy [48]. Event rates associated with exposure time can be calculated for the overall cohort and in patient subgroups. Defining exposure periods can be challenging due to the poorly understood assumptions regarding the amount of dispensed opioid, the intended and recorded days' supply, and actual patient use. As in other pharmacoepidemiologic studies, these analyses are ideally conducted in incident user cohorts. If available, validated codes or code-based claims algorithms should be used (e.g., for overdose), with portability assessment in the database being used for the study, and linkage to a source of data that captures out-of-hospital overdose deaths. Otherwise, validation should be conducted using medical records, although the lack of a true gold standard for some abuse-related outcomes can complicate these efforts. When evaluating trends over time, the effect of the transition from ICD-9 to ICD-10 must also be carefully considered [49]. Prospective cohort studies using a mix of administrative claims and in-person assessments are also possible, where dispensing claims data can be used to identify eligible patients. In these mixed-methods studies, baseline and follow-up evaluations of misuse, abuse, and addiction can be conducted using validated self- or clinician-administered instruments [50]. A limitation of these studies is that they are designed to evaluate the risk of, and risk factors for, abuse-related adverse outcomes among persons dispensed these drugs, and not in those who obtain the drugs from another source.

As noted in the previous section, drug utilization data alone are sometimes used to try to understand and compare abuse of prescription drugs in the postmarket setting. When concurrent use of two or more drug classes raises concerns about increased risks of abuse, addiction, or overdose (e.g., opioid analgesic combined with benzodiazepines), concomitancy analyses

can be conducted in drug utilization databases to understand the extent and characteristics of this potentially problematic concomitant use. "Doctor/pharmacy shopping" analyses typically use a large retail dispensing database or PDMP to create classification schemes that incorporate the number of different prescribers and pharmacies identified in overlapping or nonoverlapping controlled substance prescription dispensings [51,52]. However, these metrics have not been validated sufficiently to determine how well they distinguish between abuse-related behaviors and other drivers for these dispensing patterns that are unrelated to abuse. Research is ongoing to better understand these "doctor/pharmacy shopping" measures and the degree to which they might be useful in discriminating between therapeutic use and abuse- or addiction-related behavior [53]. Although these types of analyses can be useful descriptive and hypothesis-generating exercises, drug utilization analyses without linked outcome data are generally not appropriate for the formal assessment of abuse, as many factors unrelated to abuse may contribute to aberrant-looking drug prescribing or dispensing patterns.

Analytic Studies

More complex pharmacoepidemiologic methods are generally necessary when the objective of a study is to formally assess risk factors for abuse-related outcomes [54], or to compare abuse risk across products or time periods, for example to understand the effect of an intervention designed to reduce the risk of abuse and related adverse outcomes. As risk is a longitudinal concept, cohort studies are, in theory, ideally suited to answering these types of questions. For relatively rarer outcomes, such as fatal overdose, case–control designs may have some utility as well. Prospective cohort studies may improve ascertainment of behavioral outcomes, such as misuse and abuse, that are not captured well in claims or other electronic healthcare data. However, these may often be infeasible due to the challenges in recruiting and retaining appropriate study populations and the long follow-up that may be necessary to observe outcomes such as addiction. Designing longitudinal studies is complicated by the aforementioned challenges related to both exposure and outcome ascertainment. In addition, confounders and effect modifiers of the drug exposure–abuse–overdose causal pathway for opioids and other drugs have not been fully characterized, and are topics in need of empiric research. Selection of comparators and use of statistical applications such as propensity score matching, as well as inverse probability of treatment weighting, are all topics for further methodologic work in this area. The use of historical time periods as comparison groups (e.g., comparing the risk of overdose associated with a product before and after an intervention) is complicated by the evolving landscape of abuse trends in general, as well as insurance access and inclusion of higher- and lower-risk individuals.

When cohort studies are not feasible or warranted for a given study question, other study designs and types of data (e.g., from PCCs, ED records, and surveys) are often employed for comparative evaluation of the risk of abuse and related outcomes across multiple products. The limitations of each of these types of data must be carefully considered in the design, analysis, and interpretation of studies. Sensitivity analyses are often necessary to explore the potential effect of untestable assumptions, particularly around missing data, differential and nondifferential misclassification, and sampling bias, particularly in nonprobability samples. When appropriate for the question being posed, the prescribed availability of the drugs under study (e.g., prescriptions or tablets dispensed) may be accounted for either as an offset or as a predictor variable in statistical models. Evaluation of new market entrants and drugs with persistently small prescription volume is particularly challenging, due to low levels of population

exposure and small event counts in most abuse surveillance systems [55].

When evaluating interventions, isolating the effect of a specific intervention from the effects of other interventions and secular trends – confounders by calendar time – is a particularly challenging endeavor, especially in the case of opioid analgesics, and it is important to consider other factors that can affect prescribing and abuse trends. Examples include changes in formularies and insurance coverage policies, provider education initiatives and practice guidelines, increasing use of PDMPs, law enforcement operations to reduce inappropriate prescribing and dispensing (e.g., "pill mill" crackdowns), and the availability and cost of alternative drugs of abuse, including illicit drugs. In some instances, comparator products expected to be similarly affected by these external forces are used to control for the effects of these confounders; however, additional strategies may sometimes be warranted, for example adjusting for geographic differences in the implementation of important concurrent interventions. In most cases, one or more comparators are selected to approximate the counterfactual [56], using them to approximate the expected abuse trends and patterns in the absence of an intervention (e.g., formulation of an opioid product with abuse-deterrent properties).

When studying changes in outcomes before and after a reformulation of a drug or other intervention using a pre–post design, comparators are often incorporated into models using a drug-by-time interaction to conduct difference-in-difference analyses. One approach is to compare the mean rate or prevalence of abuse between a specified pre-period and post-period, and then to compare the magnitude of any observed change to that of a comparator; however, these analyses do not fully take into consideration preexisting secular trends, either for the drug of interest or for the comparator. Particularly in situations where baseline use and abuse rates are not stable, interrupted time-series analyses (e.g., segmented or piecewise linear regression) can be used. The latter approach is generally preferable, but is not feasible without an adequate sample size and a sufficient number of data points in each time period. These analyses also have their own set of assumptions, for example about linearity and the lack of concurrent interventions. Multiple comparators are often used because no single comparator is ideal, given inherent differences in baseline characteristics such as abuse rates and patterns. The use of multiple comparators complicates hypothesis testing and overall interpretation of findings, however. Composite comparators (multiple drugs grouped together) are sometimes used, but these may include drugs that vary widely in market share and utilization and abuse trends, further complicating interpretation. In some instances, it may be useful to include both "positive" and "negative" controls; for example, a drug known to have a high risk of abuse or expected to be greatly affected by an intervention, and another with no or low abuse potential or not expected to be affected by an intervention.

Interpretation

The interpretation of abuse-related findings depends on the study question, the source of data, and the methods used to analyze them. A review of spontaneous or anecdotal reports of abuse or abuse-related adverse outcomes may be adequate to establish a safety signal, but would not be an appropriate basis for quantifying differential risk between one product and another. Descriptive analyses of population data – for example from national surveys, ED visit data, or death certificates – are important for guiding public health and law enforcement policy, resource allocation, drug regulation, and design of interventions. Understanding the data and their limitations is critical to proper interpretation of study findings, for instance avoiding conflation of utilization and abuse trends, and

accounting for jurisdictional variation in the documentation of specific drugs in fatal drug overdoses [57].

When interpreting the results of hypothesis-testing studies, the goal is generally causal inference, determining, for example, the causal role of opioid analgesic dose in overdose risk, or the effect of an abuse-deterrent formulation on the likelihood of a drug being insufflated or injected. Often, a mosaic of different types of data and study populations must be qualitatively synthesized to make a causal inference. Fundamental epidemiologic principles such as the Bradford Hill causality criteria [58] can be helpful in this endeavor. Particularly important elements in this area are temporality, strength of association, consistency of findings across multiple populations and study designs, coherence with experimental data, specificity of effect, and consideration of alternative explanations. Ultimately, the interpretation may be a largely qualitative determination, or include a range of possible effect estimates that take into consideration both random and systematic errors and effect modifiers.

Finally, the unit of analysis and the potential for ecological fallacy must be considered. In particular, one must exercise great caution when drawing inferences about individual-level risk from ecologic analyses. For example, a study may show a decline in the prevalence of nonoral abuse of a product within a given population after that product was reformulated with properties designed to deter intranasal and injection abuse. Although a change in the prevalence of a specific drug's abuse by more dangerous routes in the community may be an important finding, one cannot directly infer that individuals exposed to the drug have a reduced risk of abusing the drug through these routes. Similarly, correlations in trends for different drug classes – for example, prescription opioid overdose deaths plateauing as heroin death rates rise [59–61] – are important for generating hypotheses, but do not provide direct evidence of individuals transitioning from one drug to another. Addressing this issue directly would require a longitudinal study analyzed at the individual level.

Currently Available Solutions

Despite their many challenges, pharmacoepidemiologic investigations and advances in methodology have contributed to answering many important clinical, public health, and regulatory questions involving pharmaceutical drug abuse and related adverse outcomes. Some examples include the three major types of questions addressed by research in this area.

Signal Detection

Pharmacoepidemiologic data have been used to determine whether a currently marketed drug is being abused in the community, and to identify important clinical adverse consequences of this behavior. A recent example was the regulatory action involving the OTC antidiarrheal drug loperamide. Loperamide is a peripherally acting opioid agonist that does not have significant CNS effects when used at recommended doses, and has been available without a prescription in the US for about 30 years. However, FAERS reports and multiple published case reports began to emerge in recent years, identifying serious cardiac arrhythmias associated with loperamide ingestion at many times the maximum recommended dose, typically among individuals attempting to self-treat opioid withdrawal symptoms or attain a euphoric effect [62]. These reports were accompanied by a published analysis of an internet discussion forum describing misuse and abuse of high doses of loperamide, as well as the reasons and subjective experiences that individuals reported [63]. This analysis noted a recent sharp increase in internet postings related to the misuse and abuse of loperamide, as well as mentions of

concomitant use of other medications in an attempt to facilitate the CNS penetration. Finally, rates of loperamide intentional exposure calls to PCCs in the US also increased markedly [64]. Based on this body of evidence, in June 2016 the FDA issued a Drug Safety Communication warning that taking higher than recommended doses of loperamide in the context of abuse or misuse of the product, or otherwise, can cause serious cardiac problems that can lead to death. In addition, the FDA required labeling changes for loperamide products to alert healthcare providers and consumers to these risks [65].

Descriptive Quantitative Assessment

Descriptive analyses of population data on drug misuse, abuse, addiction, and overdose are widely used for public health surveillance [2]. They are also used to inform decisions on drug scheduling and other regulations intended to minimize these risks. In October 2014, based on the FDA's review of the available epidemiologic and other scientific data, the Drug Enforcement Administration rescheduled hydrocodone combination products from Schedule III to the more restrictive Schedule II of the CSA. The FDA's review of the available epidemiologic data included analyses of ED visit data from the now defunct DAWN system, survey data from NSDUH and MTF, and selected state medical examiner data on drug-involved mortality. The review found that, after adjustment for prescription volume, the abuse prevalence of hydrocodone combination products appeared to be lower than that of Schedule II comparator products. However, both use *and* abuse were found to be widespread, and, from a public health standpoint, the products were determined to pose a significant absolute risk to the community [66]. Therefore, the decision to recommend the more restrictive scheduling was based largely on analyses using population-based (rather than utilization-based) denominators, which suggested a large public health burden associated with these products.

Descriptive Quantitative Assessment and Analytic Studies

Sometimes, descriptive and analytic pharmacoepidemiologic investigations are used together to guide public health or regulatory actions. A recent example is the issuance of a boxed warning – also commonly referred to as a "black box warning," which appears on a prescription drug's label and is designed to call attention to serious or life-threatening risks – describing the serious risks and death associated with concomitant use of opioids and benzodiazepines and other CNS depressants [67]. This regulatory decision relied on studies suggesting both increasing trends in concomitant prescribing, misuse, and overdoses involving opioids and benzodiazepines together, as well as an elevated risk of fatal overdose associated with concomitant use of these two drug classes. One descriptive study examined concomitant use patterns of opioid analgesics and benzodiazepines using a proprietary source of retail pharmacy dispensing data (IMS Health®). Investigators observed that despite multiple published clinical guidelines advising against co-prescribing of opioid analgesics and benzodiazepines, the proportion of opioid analgesic recipients receiving an overlapping benzodiazepine prescription increased by 41% between 2002 and 2014. They also found that a majority of these overlapping prescriptions were written by the same prescriber [68]. Another descriptive study used DAWN to analyze ED visits due to nonmedical use of both prescription opioid analgesics and benzodiazepines, and the NVSS Multiple Cause of Death file to analyze overdose deaths involving both opioid analgesics and benzodiazepines. Between 2004 and 2011, rates of both nonmedical use-related ED visits and fatal overdoses involving both drug classes

increased sharply, and the proportion of overdose deaths due to prescription opioid analgesics that also involved benzodiazepines rose from 18% to 31% [69]. Two analytic epidemiologic studies published in the medical literature suggested an increased risk of fatal overdose in patients dispensed both opioid analgesics and benzodiazepines. First, a cohort study linking North Carolina medical examiner data and state prescription drug monitoring program dispensing data found that rates of overdose death among patients co-dispensed opioid analgesics and benzodiazepines were 10 times higher than those among patients dispensed opioid analgesics alone [70]. Although the ability to control for confounding is limited in this type of study, the magnitude of effect suggested an elevated risk of death with this drug combination. Second, a case–cohort study examined Veterans Health Administration data from 2004 to 2009 and found that the risk of death from drug overdose significantly increased among patients with concomitant opioid analgesic and benzodiazepine prescriptions compared to patients taking opioid analgesics without receipt of a benzodiazepine, after controlling for both baseline and time-varying confounders [71]. In summary, a variety of pharmacoepidemiologic studies, both descriptive and analytic, were used to characterize a growing public health problem in this area and support new warnings on product labels.

Researchers, public health practitioners, policymakers, and regulators frequently use epidemiologic data to evaluate the impact of interventions intended to increase the safety of controlled substances. Much of this work in recent years has focused on opioid analgesics, for example evaluating the impact of interventions such as PDMPs [72], implementation of opioid analgesic prescribing guidelines [73], and pain clinic regulation [33]. Other subjects of interest are the assessment of opioid analgesic REMS and the effectiveness of abuse-deterrent opioid analgesic formulations. As already described, this is one of the most challenging areas of pharmaceutical drug abuse pharmacoepidemiology. The use of multiple different study designs, data sources, denominators, and comparators, as well as sensitivity analyses, can help to evaluate assumptions that are difficult to test empirically; for example, misclassification of products in self-reported data, sampling bias, the best approximation of the counterfactual, or the relationship between the prescribed availability of a drug and the likelihood of abuse in a population.

A recent example where this approach was useful was the evaluation of postmarketing data for reformulated Opana® ER (oxymorphone extended-release) [74], which ultimately resulted in removal of this product from the market [75]. Opana ER was an opioid product initially approved by the FDA in 2006. In 2012, a reformulated version was introduced with a proprietary polyethylene oxide–containing matrix designed to make the drug more difficult to crush into a powder or dissolve in solution for abuse via nonoral routes. In 2017, the FDA convened an advisory committee meeting to discuss the postmarketing safety data on the reformulated product. Key to this evaluation were epidemiologic data collected from individuals entering or being assessed for substance use disorder treatment within the National Addictions Vigilance Intervention and Prevention Program (NAVIPPRO®) Addiction Severity Index-Multimedia Version (ASI-MV®) surveillance network. These data suggested that after Opana ER was reformulated, the predominant route of abuse shifted from the intranasal to the more dangerous injection route. However, substantial changes also occurred in the geographic distribution of the sites participating in the NAVIPPRO network and in the distribution of treatment settings in which individuals were being assessed. These changes resulted in a shift in the study population toward sites, specifically in Tennessee, where both Opana ER abuse and more severe substance use disorders and injection drug use

were highly prevalent. Sensitivity analyses were used to explore this potential bias, restricting the sample to sites contributing data in every quarter of the study period, as well as stratifying analyses by site location within and outside Tennessee. The shift in the route of abuse profile for Opana ER was consistently observed in all of these analyses. PCC call data also indicated a shift in the route of abuse profile of Opana ER, although both the PCC and treatment center survey data were limited by difficulties distinguishing between brand and generic versions and original and reformulated versions of a product, a critical distinction because only the branded product was reformulated. The potential for differential misclassification across these products limited the ability to make inferences about changes in population- and utilization-based abuse rates over time. Therefore, changes in route of abuse patterns over time in the subgroup reporting abuse of Opana ER became a critical piece of information, as product misclassification is less likely to bias change estimates away from the null for these conditional probabilities.

The Opana ER example also illustrates how multiple studies, each with limitations, can be qualitatively synthesized using fundamental epidemiologic principles, including Hill's criteria, to make a determination regarding the effect of an intervention. The marked shift from intranasal to injection abuse of Opana ER corresponded temporally with initial marketing of the reformulated product, and a shift of similar magnitude was not observed for any other opioid comparator analyzed. The shift in the route of abuse was observed in multiple epidemiologic data sources and populations, and was coherent with other types of data, including premarket experimental abuse liability data. The patterns seen in the epidemiologic data were also consistent with spontaneous and anecdotal reports. A FAERS case series indicated that prior to Opana ER's reformulation, the vast majority of reported cases of nonoral abuse described the intranasal route, whereas after the reformulation, cases of intravenous abuse began to emerge. The review of FAERS reports also identified an association between intravenous use of reformulated Opana ER and a rare hemolytic microangiopathy. Finally, investigation of an unprecedented HIV outbreak in rural Indiana indicated that Opana ER abuse via injection was virtually universal among those newly infected with HIV, and that users reported switching from the intranasal to the injection route after the drug was reformulated. Interviews with drug users in the community indicated that the hardening and gelling properties of reformulated Opana ER were driving users to increase the volume of solvent and to share pills and "cooking" equipment, resulting in increased injections and opportunities for transmission of blood-borne infections [76]. In summary, although each individual study and source of postmarketing data had limitations, synthesis of all the information using fundamental principles of causal inference led to a determination that the benefits of reformulated Opana ER as an opioid analgesic option no longer outweighed the risks associated with abuse of this product.

The Future

The opioid crisis has elevated the urgency of pharmacoepidemiologic research on drugs of abuse; however, data and methods developed to improve research on opioid analgesics are valuable for studying abuse of other pharmaceuticals as well. Robust surveillance systems and systematic data linkages are needed to support these investigations, and efforts are underway to develop such data systems at the national and state levels. Although no single data source or method is ever expected to be sufficient to address all questions about drug abuse, the most useful data systems and methods to support

future pharmacoepidemiologic research in this area would feature several characteristics:

- Identification of specific products, formulations, and routes of abuse, with the ability to evolve in response to changes in the pharmaceutical market.
- Rigorous development and validation of survey instruments and coded algorithms, using innovative methods to minimize misclassification.
- Use of probability sampling methods when feasible. Where infeasible (e.g., when using emerging approaches such as internet/mobile phone surveys), conducting external validation studies to ensure that results can be applied to a well-defined target population that is stable over time.

Further empiric research is needed to identify and appropriately adjust for individual- and community-level confounding and to better characterize the relationship between prescription volume and abuse rates, in order to understand the degree to which prescribing is driven by the demand of those intending to abuse or divert the drug. More studies using longitudinal designs are also needed to understand the progression from therapeutic use of pharmaceutical products to abuse, addiction, and related harms, and where we can intervene to prevent these outcomes.

Finally, the opioid crisis has highlighted the intertwined nature of pharmaceutical and illicit drug abuse. This interconnection emphasizes the sociocultural and economic forces driving drug abuse, and has spurred interest in methods that fall outside of traditional pharmacoepidemiology but are well established in the broader field of substance abuse research. These often include more qualitative and mixed-methods approaches that may enhance our understanding of patterns observed in the epidemiologic data.

References

1 Centers for Disease Control and Prevention. Opioid overdose: understanding the epidemic. 2018. https://www.cdc.gov/drugoverdose/epidemic/index.html (accessed November 2017).

2 Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Annual surveillance report of drug-related risks and outcomes: United States, 2017. https://www.cdc.gov/drugoverdose/pdf/pubs/2017-cdc-drug-surveillance-report.pdf (accessed November 2017).

3 U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Abuse-deterrent opioids – evaluation and labeling: guidance for industry. April 2015. https://www.fda.gov/downloads/Drugs/Guidances/UCM334743.pdf (accessed November 2017).

4 Substance Abuse and Mental Health Services Administration. Reports and detailed tables from the 2016 National Survey on Drug Use and Health (NSDUH). https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf (accessed May 2019).

5 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Washington, DC: American Psychiatric Association, 2013.

6 Embeda (morphine sulfate and naltrexone hydrochloride extended-release capsules) approved package insert, Section 9.2: Abuse. https://www.accessdata.fda.gov/drugsatfda_docs/label/2016/022321s022lbl.pdf (accessed November 2017).

7 US Drug Enforcement Administration. Chapter 1: The Controlled Substances Act. https://www.dea.gov/controlled-substances-act (accessed November 2017).

8 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Assessment of abuse potential of drugs: guidance for industry. January 2017. https://www.fda.gov/downloads/drugs/guidances/ucm198650.pdf (accessed November 2017).

9 Warner M, Paulozzi LJ, Nolte KB, Davis GG, Nelson LS. State variation in certifying manner of death and drugs involved in drug intoxication deaths. *Acad Forensic Pathol* 2013; **3**: 231–7.

10 Toxicology Investigators Consortium (ToxIC). About ToxIC. http://www.toxicregistry.org (accessed November 2017).

11 Researched Abuse, Diversion and Addiction-Related Surveillance (RADARS) System. Web monitoring. 2019. https://www.radars.org/radars-system-programs/web-monitoring.html (accessed May 2019).

12 Inflexxion. Health research and analysis. 2017. https://www.inflexxion.com/health-research-analysis (accessed November 2017).

13 Researched Abuse, Diversion and Addiction-Related Surveillance (RADARS) System. StreetRx. 2019. https://www.radars.org/radars-system-programs/streetrx.html (accessed May 2019).

14 Researched Abuse, Diversion and Addiction-Related Surveillance (RADARS) System. Drug Diversion. 2019. https://www.radars.org/radars-system-programs/drug-diversion.html (accessed May 2019).

15 Johnston LD, O'Malley PM, Miech RA, Bachman JG, Schulenberg JE. Monitoring the future: national survey results on drug use 1975–2016. 2016 overview: key findings on adolescent drug use. Monitoring the Future Project. 2016. http://www.monitoringthefuture.org/pubs/monographs/mtf-overview2016.pdf (accessed November 2017).

16 Arria AM, Caldeira KM, O'Grady KE, Vincent KB, Johnson EP, Wish ED. Nonmedical use of prescription stimulants among college students: associations with attention-deficit-hyperactivity disorder and polydrug use. *Pharmacotherapy* 2008; **28**:156–69.

17 Cassidy TA, Varughese S, Russo L, Budman SH, Eaton TA, Butler SF. Nonmedical use and diversion of ADHD stimulants among U.S. adults ages 18–49: a national internet survey. *J Atten Disord* 2015; **19**: 630–40.

18 RADARS System. Prevalence of illicit drug use: survey of non-medical use of prescription drugs program compared to NSDUH [abstract]. College on Problems of Drug Dependence Annual Meeting, June 2017. https://www.radars.org/system/publications/CPDD%202017%20NMURX%20US%20vs%20NSDUH%20poster.pdf (accessed May 2019).

19 Substance Abuse and Mental Health Services Administration. Treatment Episode Data Set (TEDS). 2017. https://www.samhsa.gov/data/client-level-data-teds (accessed November 2017).

20 Butler SF, Budman SH, Licari A, *et al*. National addictions vigilance intervention and prevention program (NAVIPPRO): a real-time, product-specific, public health surveillance system for monitoring prescription drug abuse. *Pharmacoepidemiol Drug Saf* 2008; **17**: 1142–54.

21 Researched Abuse, Diversion and Addiction-Related Surveillance (RADARS) System. Opioid Treatment and Survey of Key Informants' Patients. 2019. https://www.radars.org/radars-system-programs/opioid-treatment.html and https://www.radars.org/radars-system-programs/survey-of-key-informants-patients.html (accessed May 2019).

22 Cassidy TA, DasMahapatra P, Black RA, Wieman MS, Butler SF. Changes in prevalence of prescription opioid abuse after introduction of an abuse-deterrent opioid formulation. *Pain Med* 2014; **15**: 440–51 (Table 4).

23 By K, McAninch JK, Keeton SL, *et al*. Important statistical considerations in the evaluation of post-market studies to assess whether opioids with abuse-deterrent properties result in reduced abuse in the

community. *Pharmacoepidemiol Drug Saf* 2017; **27**(5). doi:10.1002/pds.4287

24 American Association of Poison Control Centers. National Poison Data System (NPDS). 2017. http://www.aapcc.org/data-system/ (accessed November 2017).

25 Dasgupta N, Davis J, Jonsson Funk M, Dart R. Using poison center exposure calls to predict methadone poisoning deaths. *PLoS One* 2012; **7**: e41181.

26 Iwanicki JL, Severtson SG, Margolin Z, Dasgupta N, Green JL, Dart RC. Consistency between opioid-related mortality trends derived from Poison Center and National Vital Statistics System, United States, 2006–2016. *Am J Public Health* 2018; **108**(12): 1639–45.

27 Substance Abuse and Mental Health Services Administration. Emergency department data. 2017. https://datafiles.samhsa.gov/study/drug-abuse-warning-network-dawn-2011-nid13586 (accessed May 2019).

28 Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). 2016. https://www.ahrq.gov/research/data/hcup/index.html (accessed November 2017)

29 Chou R, Turner JA, Devine EB, *et al*. The effectiveness and risks of long-term opioid therapy for chronic pain: a systematic review for a National Institutes of Health Pathways to Prevention Workshop. *Ann Int Med* 2015; **162**: 276–86.

30 Cochran G, Woo B, Lo-Ciganic W-H, Gordon AJ, Donohue JM, Gellad WF. Defining non-medical use of prescription opioids within health care claims: a systematic review. *Subst Abus* 2015; **36**: 192–202.

31 Canan C, Polinski JM, Alexander GC, Kowal MK, Brennan TA, Shrank WH. Automatable algorithms to identify nonmedical opioid use using electronic data: a systematic review. *J Am Medical Inform Assoc* 2017; **24**:1204–10.

32 Centers for Disease Control and Prevention, National Center for Health Statistics. National Death Index. 2017. https://www.cdc.gov/nchs/ndi/index.htm (accessed November 2017).

33 Johnson H, Paulozzi L, Porucznik C, Mack K, Herter B. Decline in drug overdose deaths after state policy changes—Florida, 2010–2012. *Morb Mortal Wkly Rep* 2014; **63**: 569–74.

34 Marteau D, McDonald R, Patel K. The relative risk of fatal poisoning by methadone or buprenorphine within the wider population of England and Wales. *BMJ Open* 2015; **5**: e007629.

35 Centers for Disease Control and Prevention. ICD injury matrices. 2017. https://www.cdc.gov/nchs/injury/injury_matrices.htm (accessed November 2017).

36 Centers for Disease Control and Prevention. Drug Induced Mortality (DIM) restricted variable. 2017. https://www.cdc.gov/rdc/b1datatype/dt1229.html (accessed November 2017).

37 Trinidad JP, Warner M, Bastian BA, Minino AM, Hedegaard H. Using literal text from the death certificate to enhance mortality statistics: characterizing drug involvement in deaths. *Natl Vital Stat Rep* 2016; **65**: 1–15.

38 Centers for Disease Control and Prevention. Enhanced state opioid overdose surveillance. 2017. https://www.cdc.gov/drugoverdose/foa/state-opioid-mm.html (accessed November 2017).

39 US Department of Justice, Drug Enforcement Administration Diversion Control Division. National Forensic Laboratory Information System (NFLIS). 2017. https://www.deadiversion.usdoj.gov/nflis (accessed November 2017).

40 Fuzier R, Serres I, Guitton E, Lapeyre-Mestre M, Montastruc JL; French Network of Pharmacovigilance Centres. Adverse drug reactions to gabapentin and pregabalin: a review of the French pharmacovigilance database. *Drug Safety* 2013; **36**: 55–62.

41 Schwan S, Sundström A, Stjernberg E, Hallberg E, Hallberg P. A signal for an abuse liability for pregabalin: results from the Swedish spontaneous adverse drug reaction reporting system. *Eur J Clin Pharmacol* 2010; **66**: 947–53.

42 Bossard JB, Ponté C, Dupouy J, Lapeyre-Mestre M, Jouanjus E. Disproportionality analysis for the assessment of abuse and dependence potential of pregabalin in the French pharmacovigilance database. *Clin Drug Investig* 2016; **36**: 735–42.

43 McNaughton EC, Black RA, Zulueta MG, Budman SH, Butler SF. Measuring online endorsement of prescription opioids abuse: an integrative methodology. *Pharmacoepidemiol Drug Saf* 2012; **21**:1081–92.

44 McNaughton EC, Coplan PM, Black R, Weber SE, Chilcoat HD, Butler, SF. Monitoring of internet forums to evaluate reactions to the introduction of reformulated OxyContin to deter abuse. *J Med Internet Res* 2014; **16**: e119.

45 National Cancer Institute. Joinpoint Regression Program: average annual percent change. 2014. https://surveillance.cancer.gov/joinpoint (accessed May 2019).

46 Secora A, Trinidad JP, Zhang R, Gill R, Dal Pan G. Drug availability adjustments in population-based studies of prescription opioid abuse. *Pharmacoepidemiol Drug Saf* 2017; **26**: 180–91.

47 Cassidy TA, Oyedele N, Mickle TC, Guenther S, Budman SH. Patterns of abuse and routes of administration for immediate-release hydrocodone combination products. *Pharmacoepidemiol Drug Saf* 2017; **26**: 1071–82.

48 FDA Letter to Application Holders—ER/LA Opioid Analgesic Class. Postmarketing Requirement #2065 to #3033 Release and Reissue. Page 4: PMR 3033-2. 2017. https://www.fda.gov/downloads/Drugs/DrugSafety/InformationbyDrugClass/UCM484415.pdf (accessed November 2017).

49 Heslin KC. Owens PL, Karaca Z, Barrett ML, Moore BJ, Elixhauser Z. Trends in opioid-related inpatient stays shifted after the US transitioned to ICD-10-CM diagnosis coding in 2015. *Med Care* 2017; **55**: 918–23.

50 FDA Letter to Application Holders—ER/LA Opioid Analgesic Class. Postmarketing Requirement #2065 to #3033 Release and Reissue. Page 4: PMR 3033-1. 2017. https://www.fda.gov/downloads/Drugs/DrugSafety/InformationbyDrugClass/UCM484415.pdf (accessed November 2017).

51 Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. *Drug Saf* 2012; **35**: 325–34.

52 Han H, Kass PH, Wilsey BL, Li C-S. Increasing trends in Schedule II opioid use and doctor shopping during 1999–2007 in California. *Pharmacoepidemiol Drug Saf* 2014; **23**: 26–35.

53 FDA Letter to Application Holders—ER/LA Opioid Analgesic Class. Postmarketing Requirement #2065 to #3033 Release and Reissue. Page 6-7: PMR 3033-8, 3033-9, and 3033-10. 2017. https://www.fda.gov/downloads/Drugs/DrugSafety/InformationbyDrugClass/UCM484415.pdf (accessed November 2017).

54 Bohnert AS, Valenstein M, Bair MJ, *et al*. Association between opioid prescribing patterns and opioid overdose-related deaths. *JAMA* 2011; **305**: 1315–21.

55 Roland CL, Wetnick B, Brown DA. Assessing the impact of abuse-deterrent opioids (ADOs): identifying epidemiologic factors related to new entrants with low population exposure. *Postgrad Med* 2017; **129**: 12–21.

56 Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health* 2001; **22**: 189–212.

57 Ruhm CJ. Geographic variation in opioid and heroin involved drug poisoning mortality rates. *Am J Prev Med* 2017; **53**: 745–53.

58 Hill AB. The environment and disease: association or causation? *J Royal Society Med* 1965; **8**: 295–300.

59 Rudd RA, Seth P, David F, Scholl L. Increases in drug and opioid-involved overdose

deaths - United States, 2010–2015. *Morb Mortal Wkly Rep* 2016; **65**: 1445–52.

60 Unick GJ, Ciccarone D. US regional and demographic differences in prescription opioid and heroin-related overdose hospitalizations. *Int J Drug Policy* 2017; **46**: 112–19.

61 Compton WM, Jones CM, Baldwin GT. Relationship between nonmedical prescription opioid use and heroin use. *N Engl J Med* 2016; **374**: 154–63.

62 Swank KA, Wu E, Kortepeter C, McAninch J, Levin RL. Adverse event detection using the FDA post-marketing drug safety surveillance system: cardiotoxicity associated with loperamide abuse and misuse. *J Am Pharm Assoc* 2017; **57**: S63–S67.

63 Daniulaitye R, Carlson R, Falck R, *et al*. "I just wanted to tell you that loperamide will work": a web-based study of extra-medical uses of loperamide. *Drug Alcohol Depend* 2013; **130**: 241–4.

64 American Association of Poison Control Centers. Annual reports of the American Association of Poison Control Centers' (AAPCC) National Poison Data System (NPDS), years 2006–2014. http://www.aapcc.org/annual-reports (accessed November 2017).

65 FDA Drug Safety Communication: FDA warns about serious heart problems with high doses of the antidiarrheal medicine loperamide (Imodium), including from abuse and misuse. https://www.fda.gov/Drugs/DrugSafety/ucm504617.htm (accessed November 2017).

66 U.S. Department of Justice, Drug Enforcement Administration. 21 CFR Part 1308 [Docket No. DEA-389] Schedules of Controlled Substances: Rescheduling of Hydrocodone Combination Products from Schedule III to Schedule II. 2014. https://www.deadiversion.usdoj.gov/fed_regs/rules/2014/fr0822.htm (accessed November 2017).

67 FDA Drug Safety Communication: FDA warns about serious risks and death when combining opioid pain or cough medicines with benzodiazepines; requires its strongest warning. 2017. https://www.fda.gov/Drugs/DrugSafety/ucm518473.htm (accessed November 2017).

68 Hwang CS, Kang EM, Kornegay CJ, Staffa JA, Jones CM, McAninch JK. Trends in the concomitant prescribing of opioids and benzodiazepines, 2002–2014. *Am J Prev Med* 2016; **51**: 151–60.

69 Jones CM, McAninch JK. Emergency department visits and overdose deaths from combined use of opioids and benzodiazepines. *Am J Prev Med* 2015; **49**: 493–501.

70 Dasgupta N, Funk MJ, Proescholdbell S, Hirsch A, Ribisl KM, Marshall S. Cohort study of the impact of high-dose opioid analgesics on overdose mortality. *Pain Med* 2016; **17**: 85–98.

71 Park TW, Saitz R, Ganoczy D, Ilgen MA, Bohnert AS. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics: case–cohort study. *BMJ* 2015; **350**: h2698.

72 Prescription Drug Monitoring Program Center of Excellence at Brandeis. Briefing on PDMP effectiveness. September 2014. http://www.pdmpassist.org/pdf/COE_documents/Add_to_TTAC/Briefing%20on%20PDMP%20Effectiveness%203rd%20revision.pdf

73 Von Korff M, Walker RL, Saunders K, *et al*. Prevalence of prescription opioid use disorder among chronic opioid therapy after health plan opioid dose and risk reduction initiative. *Int J Drug Policy* 2017; **46**: 90–98.

74 FDA Advisory Committee Briefing Document. Joint Meeting of the Drug Safety and Risk Management (DSaRM) Advisory Committee and the Anesthetic and Analgesic Drug Products Advisory Committee (AADPAC). March 13-14, 2017. Postmarketing safety issues

related to reformulated Opana® ER. https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AnestheticAndAnalgesicDrugProductsAdvisoryCommittee/UCM545760.pdf (accessed November 2017).

75 FDA News Release: FDA requests removal of Opana® ER for risks related to abuse. 2017. https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm562401.htm (accessed November 2017).

76 Brooks JT. CDC outbreak investigations involving Opana® ER. Slide presentation for the Joint Meeting of the Drug Safety and Risk Management Advisory Committee and the Anesthetic and Analgesic Drug Products Advisory Committee. March 13-14, 2017. https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AnestheticAndAnalgesicDrugProductsAdvisoryCommittee/UCM547237.pdf

Part V

Selected Special Methodologic Issues in Pharmacoepidemiology

29

Assessing Causation from Case Reports

Judith K. Jones[1], Bernard Bégaud[2], and Elyse Kingery[3]

[1] *PharmaLex, Inc., Fairfax, VA; University of Michigan School of Public Health Summer Program, Ann Arbor, MI, USA*
[2] *University of Bordeaux School of Medicine, Bordeaux, France*
[3] *Degge Group, Fairfax, VA, USA*

The evaluation of reports of suspected adverse drug reactions in the clinical setting or in a clinical trial is a judgment about whether – and the degree to which – any reported event is, in fact, causally associated with one or more suspected drug(s). In reality, a particular event either *is* or *is not* associated with a particular product. However, our current tools almost never allow a definitive determination. Accordingly, a number of approaches to the assessment of the probability of a causal drug–event association have evolved and are utilized in determining the possibility that a drug contributed to an event. This chapter will discuss the evolution of these efforts and several of the current approaches and uses. It will then review the evolving regulatory changes on this topic, including a brief consideration of the evaluation of single events in the clinical trial setting. The focus of this chapter is on causal assessment of individual events and differs from the approach to assessing general causality as discussed in Chapter 3.

Understanding Potential Causal Relationships Supports Clinical and Regulatory Evaluations and Actions

Figure 29.1 illustrates the basic clinical problem to be addressed: An adverse event occurs within the milieu of a number of possible causal factors. That event either occurred independently or was partially or totally linked to one or more of the potential causative agents. The assessor's task is to determine the degree to which the occurrence of the event is linked to one particular suspected causal agent – a drug or other medical product.

This task is similar to evaluating causality in chronic disease epidemiology, as discussed below. However, in disease epidemiology, causality relates to events in populations and to the assessments of those events in one or more population studies. In medical products

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

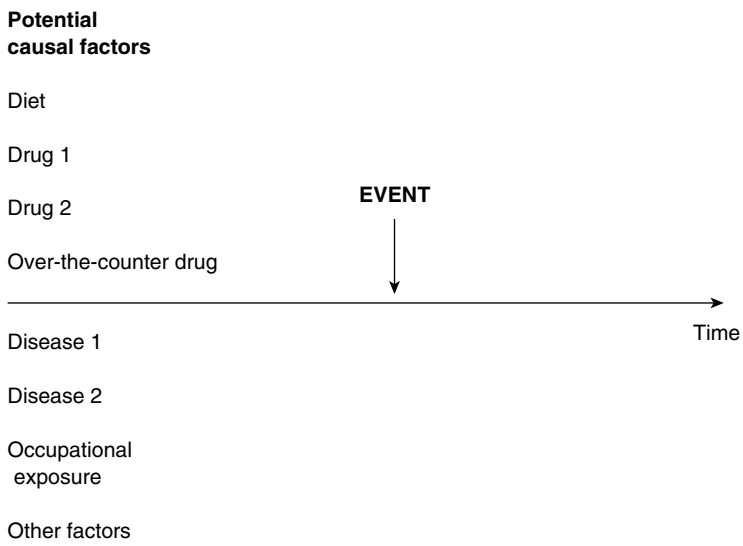


Figure 29.1 Diagram depicting the dilemma for determining causation of an event in a clinical setting. In reality, a drug either did or did not cause or contribute to an event. However, given the multiple factors associated with the event, the actual truth can seldom be ascertained. Instead, some expression of probability that the drug was associated with the event is made. The method by which this expression is determined is the primary concern of those in adverse reaction causality research.

epidemiology, assessing causality from individual case reports of suspected adverse reactions to a medical product represents a different, equally complex challenge due to multiple factors.

Observations of medical events and resulting reports are often *biased*: A reporter reports because he/she suspects a causal relationship between the product and the event. Thus, alternative causes can be overlooked or ignored and few or no data are collected that might support the alternative(s).

Data are frequently *incomplete*, often omitting details on:

- reason for use
- age, gender
- product exposure including dose, actual dose ingested, duration of use, prior exposure
- route of administration
- concomitant medical products
- patient medical history
- details of the adverse event including onset, characteristics, time course, and outcome
- baseline laboratory data and other important clinical factors (because the suspicion is usually retrospective and the desired data are often not available when the report is made)
- whether the event stopped when the product was withdrawn (dechallenge)
- whether the event returned when the product was resumed (rechallenge)
- outcome of the event
- follow-up information.

Case reports from clinical trials and hospital settings tend to be more complete than reports submitted in the postmarketing setting.

Adverse reactions can be acute, subacute, or chronic, can be reversible or not (e.g., death and birth defects), can be rare or common, and can be pathologically unique or identical to known common diseases. Therefore, the challenge has been to define general data elements and criteria for assessing causality that will apply to *most* types of suspected adverse reactions. For example, for irreversible events such as birth defects or death, data on dechallenge (what occurs when a drug is discontinued) and rechallenge (what occurs when a drug is reintroduced) are irrelevant.

Since the assessor must make, at the very least, an implicit judgment of causality, and since evaluations of case reports are an important part of postmarketing surveillance, methods

were devised in an attempt to develop a coherent, consistent, and reliable way to assess the degree, if any, of a causal relationship between exposures and events.

Assessment of a potential causal relationship is closely linked to the reason for making that particular causality assessment, and the impact of that judgment. If the causality assessment is perceived to have little impact on future actions relating to either a patient in a clinical setting or to product labeling in the regulatory environment, it might logically be less rigorous. Conversely, if continuation of a clinical trial or continued approval of a drug hinges upon the assessment, the reliability of the method becomes more critical. The need for consistent and reliable methods of causality determination [1] has become more important with greater focus on the entire subject of adverse drug reactions, the introduction of concepts of causality assessment into more drug regulatory language, and product liability.

Moreover, worldwide and country/region-specific regulations all require reporting of events in clinical trials and in the postmarketing period. This chapter will discuss the US and European requirements briefly as well as the French method.

Historical Perspectives

Development of Concepts of Causality for Adverse Reactions

The development of thinking about the causality of adverse drug reactions evolved in two disciplines: epidemiology and the study of individual case reports of adverse reactions. Consideration of both is important.

In the 1950s, epidemiologists grappled with the issue of disease causality. Yerushalmy and Palmer [2] developed a set of proposed criteria for causality related to the association of exposures with events. They drew upon the Bradford Hill causality criteria (described in more detail in Chapter 3) as well as the Koch–Henle postulates for establishing causation for infectious diseases. After considerable deliberation with other epidemiologists, Yerushalmy and Palmer's method was refined into the following five criteria to determine the causal nature of an association.

1) Consistency
2) Strength
3) Specificity
4) Temporal relationship
5) Coherence, or biological plausibility [3]

While actively discussed and criticized [4], these criteria continue to be generally used in chronic disease epidemiology. They are most appropriately applied to population-based data rather than in evaluating individual cases or groups of cases from poorly defined populations. However, sometimes when large numbers of cases are considered, possibly along with population-based data on an adverse event, Yerushalmy and Palmer's criteria are invoked. For example, they form the basis for the World Health Organization's (WHO) evaluation of collective data on vaccine adverse effects by the Global Advisory Committee on Vaccine Safety of the Immunization Safety Priority Project [4]. Shakir and Layton cited these criteria as useful for considering the overall data, including spontaneous reports, on an adverse event [5]. Although seldom explicitly noted, the reasoning behind Yerushalmy and Palmer's criteria appeared at about the same time as thinking about the causal assessment of individual reports of suspected adverse drug reactions.

In the past, and occasionally currently, association between a drug and a reported adverse event was typically assumed if there were a number of similar reports [6]. Considerations of pharmacologic plausibility, dose–response, and timing factors were sometimes implicit but seldom explicit. More recently, this tendency has been supplanted by more specific methods,

proposed and in use since the 1980s, which will be detailed later in this chapter.

One of the most widely used tools is known as either "global introspection" or "unstructured clinical judgment." In it, more perplexing single or multiple suspected drug–event associations get referred to one or more experts. The experts collect all the facts relevant to the problem, compile them, and make *unstructured judgments* to assess the potential relationship. The answer is usually expressed in terms of a qualitative probability scale: "definite"; "probable"; "possible"; "doubtful"; or "unrelated" [7].

The recognized subjective nature of global introspection as an approach prompted the development of more structured methods of causality assessment. Irey, in examining the details of cases of suspected adverse reactions at the US Armed Forces Institute of Pathology in the 1960s, clearly demonstrated the discrepancy between cases initially reported as drug associated and those smaller number of cases found by careful detailed examination to likely *be* drug associated [8,9]. Shortly thereafter, clinical pharmacologists Karch and Lasagna also recognized the inadequacy of expert "global" evaluations of adverse reactions and developed an algorithm to segment the evaluation of a case into several components [10]. These two groups of investigators identified very similar basic data elements that they felt were necessary for a more standardized assessment.

- The timing of the event relative to the exposure.
- The presence or absence of other factors which might also cause the event.
- Dechallenge.
- Rechallenge.
- Other data supporting an association, for example, previous cases.

These criteria are specifically related to the special characteristics of suspected adverse drug reactions. They apply to causality assessments using either a single case or a group of cases from an ill-defined exposed population. Thus, it was thought that there was only a partial correspondence to the Bradford Hill criteria for chronic disease epidemiology; but in fact, the temporal relationship does also apply. Furthermore, in assessing the causality of either a single report or a series of cases outside an epidemiological context, there is no way to evaluate the consistency, strength or specificity of the association. The exception would be some rare, drug-associated disorders where the event in fact was uniquely and specifically (due to a defined pharmacologic or biologic mechanism) associated with a drug. For example, some patients treated with sulindac developed renal stones containing sulindac crystals.

Further, Aronson and Hauben proposed four types of adverse reactions reports where "attribution to the drug is either irrefutable or demonstrable to a high level of confidence" [11].

1) Deposition of the drug or metabolite in extracellular or intracellular tissue.
2) A very specific anatomic location or pattern of injury, such as injection site edema or inflammation.
3) Direct tissue injury or physiologic dysfunction that can be proven by physicochemical testing, such as esophageal injury with bisphosphonates.
4) Infection resulting from administration of an infectious agent, such as in bacterially contaminated injections.

Following the introduction of these assessment methods, several other approaches were developed [12–20], either as algorithms, decision tables or, in at least one case, as a diagrammatic method [19,20]. Many of these were reviewed and summarized in monographs from conferences held in the early 1980s on the causality of adverse drug reactions – two in Morges, Switzerland [21], and another in Crystal City, Virginia [22]. Most of these methods shared the basic elements originally suggested by Irey and Karch and Lasagna, but some added numerous

other details useful for the evaluation of special cases, such as injection site reactions or *in vitro* verification [18]. Some included extensive scoring systems linked to relatively extensive algorithms, such as the approach published by Kramer *et al.* [12]. Selected methods from Kramer *et al.* are discussed in detail in the next section of this chapter.

The Morges conferences [21], the 1983 Crystal City conference [22], and a 1983 Paris meeting [23] were all convened to compare a number of these approaches and to consider whether a single "gold standard" method might be developed that could represent an international consensus to be used by regulators and pharmaceutical sponsors alike. An international study group, the Active Permanent Workshop of Imputologists (APWI) ("imputology" is the French term for causality assessment), was initiated at Morges and continued into the 1990s [24–26]. Although a consensus method was not developed, a theoretical statistician, Dr David Lane, an invited outside observer at the Crystal City conference, provided an appraisal of the deliberations [27]. His critique and subsequent participation in the Paris conference and APWI resulted in the development of a new approach for assessing causality based on the Bayes probability theorem [28]. This approach considered the probability of an event occurring in the presence of a drug relative to its probability of occurring in the absence of the drug, considering all details of the case [29–32]. Although applied elsewhere in medicine, the Bayesian method had not been applied to suspected adverse effects analyses. A discussion of its application follows in the next section.

After this flurry of activity in the mid-1980s, there was more limited activity in adverse event causality, primarily marked by efforts in France in the mid-1990s. Causality assessment is a regulatory requirement in French reporting. This resulted in further elaboration of the Bayesian method by Bégaud and colleagues in Bordeaux [33] and development of a further scoring method, RUCAM, by Bénichou and Danan [34–36].

Although a standard method has not been adopted since this time, causality assessment by varying methods has diffused into other regulatory requirements in the European Union (EU), Canada, and the US, into the requirement for publication of reports in at least one journal (*Annals of Pharmacotherapy*), and, sporadically, into analyses of both clinical trial data and spontaneous reports, as described below.

Actual and Potential Uses of Causality Assessment

Despite the proliferation of methods and the great interest in adverse effects of drugs, the actual use of causality assessment methods for decision making has been infrequent; it may increase as interest in various methods of analysis of adverse events burgeons, and as old tools are adapted to new technology [37]. However, causality assessment has been required in France for many years [38] and has been formally considered in a European Community Directive [39]. This has resulted in a general consensus on the causality terms used by the EU member states [440–42]. Further, in 1994, a formal method of causality assessment for reports of vaccine-associated adverse events was instituted by Health Canada's Vaccine Safety Surveillance Section, Division of Immunization, Laboratory Center for Disease Control, conducted by the Advisory Committee on Causality Assessment [43,44].

In fact, standard assessments of causality could be useful in a variety of settings, from the clinical trials activities in drug development by the pharmaceutical manufacturer, to evaluation and monitoring of postmarketing spontaneous reports by both sponsors and regulators, to the clinical setting, where suspected adverse reactions should be a common component of the differential diagnosis, and even possibly to the courtroom and the newsroom [1].

Pharmaceutical Manufacturers

Manufacturers of biopharmaceuticals must view causality assessment for events associated with their products from the standpoints of both regulatory requirements and product liability.

The appearance of US Food and Drug Administration (FDA) regulations and draft guidances for reporting adverse events in clinical trials that are "reasonably" associated with a drug [45,46] indicates a growing need to describe the basis for defining an association that began within this setting. In September 2010, the FDA published a Draft Guidance for safety reporting in clinical trials that updated the current requirements in some respects [47]. In March 2003, the FDA proposed broadening the definition of causality to "a causal association cannot be ruled out." However, after careful consideration of the many public comments that expressed concern over what was deemed a very broad definition, in the final regulation the FDA agreed to maintain the former requirement that an event must be "reasonably associated" for it to warrant reporting [47].

Until these regulations took effect in 2011, US biopharmaceutical manufacturers did not need to consider causality assessments of adverse events in clinical trials for regulatory purposes. Regulations covering postmarketing event monitoring in the US required reporting of *all* events associated with the drug "whether or not thought to be associated with the drug" [47]. Current US FDA regulations require causality assessment for determination of reporting certain types of clinical events in clinical trials and in the Investigational New Drug (IND) regulations (21 CFR §312.22) [47], reporting of serious, unexpected events associated with use of a drug where there is a "reasonable possibility" that the events may have been caused by the drug is required. The regulations also include a disclaimer that such a report does not constitute an admission that the drug caused the event. No causality criteria or a suggested method of assessment are provided; however, they do imply that such methods might be useful.

In postmarketing regulations that became effective on June 30, 2006, the FDA modified its standard for including postmarketing safety information on the labeling [48]. The current regulatory standard for addition of an event to the product label Warnings section currently states: "The labeling must be revised to include a warning about a clinically significant hazard as soon as there is reasonable evidence of a causal association with a drug; a causal relationship need not have been definitely established" (21 CFR 201.57(c)(6)).

Outside the US, the requirements for manufacturers to consider causality varied from country to country. With promulgation of EU directives on pharmacovigilance and related EU activities, an increase in regulatory harmonization activities, and the expanded globalization of biopharmaceuticals, there are more similarities in regulatory approaches than differences. Many regulatory agencies have requested or implied some type of evaluation to minimize the number of nonspecific events reported [36]. Given this environment, particularly in a growing international milieu, manufacturers have been actively interested in this area. In fact, several of the specific methods for causality assessment have been published by investigators based in the biopharmaceutical industry [49–52].

The impact of the United Kingdom's exit from the EU in 2019 is as yet unknown. It is comforting to know that in September 2017, the UK government published a paper setting its objectives for continued scientific cooperation and collaboration with the EU, and the field of biopharmaceuticals is considered particularly important. The EMA announced in November 2017 that the new location of the organization will be in Amsterdam.

Causality definitely is an issue for pharmaceutical manufacturers in the arena of product

liability, especially in the US (see Chapter 9). In 1984, Freilich considered many aspects of this [53], concluding that a company must have a rigorous process for the review of any adverse event reports and "make causality assessments on an ongoing basis" for product liability purposes. This is necessary to comply with the duty to warn, which he summarized as follows: "Information must be given of any risks of death or serious harm, no matter how rare, as well as information concerning side effects where there is a substantial probability of their occurrence, no matter how mild." Others in the legal arena dealing in product liability have considered causality issues and the notion of the "substantial factor" test for contributing to causation [54,55]. A substantial factor is one that by itself may possibly have caused a plaintiff's injury, but that may not be the only factor involved in the injury.

Drug Regulator

Causality assessment of spontaneously reported postmarketing adverse reactions by drug regulators has varied considerably. Most countries' drug regulators have some method of approaching causality, but this has been most well defined in France, Australia, and certain other countries [49,56,57].

In France, owing in part to the considerable original work on and interest in adverse reaction causality by a regulator, J. Dangoumau, and his colleagues, all reports of suspected reactions must be evaluated by the "French method" [58]. This combines symptom and chronologic criteria relating to the individual case to give a "Global Intrinsic Score" and then adds bibliographic data relating to information on other cases and the known pharmacology and adverse effects of the drug from standardized sources to give an "Extrinsic Score" [14,58]. This method, still mandatory in France, was updated in 2011 by an *ad hoc* working group including regulators and members of pharmaceutical companies. It now provides a seven-degree score, considers the degree of completeness of data in the case, and differentiates between *expected* and *unexpected* adverse reactions [59,60].

In the US, no formal method for evaluating all reports was used until a simple algorithm was developed in the early 1980s, based on the Irey and Karch and Lasagna work [16,17]. This simple, basic method considered the timing, de- and rechallenge, and confounding factors criteria. It specifically *excluded* examining literature reports when considering the strength of the association. It was reasoned that, in many cases, the FDA would be in the position of receiving the first reports of an association. The primary use of the assessment by the FDA was administrative, as the causality assessment was a mechanism for identifying the best documented cases – those with a "probable" or "highly probable" association. The causality judgment was specifically deleted from publicly available files, which consistently carry the caveat "a causal relationship need not have been definitely established."

Use of the FDA algorithm ended in 1986, but the caveat stating that no cause–effect relationship could be derived on all released adverse event information remains. The FDA does not now use formal causality assessment on a routine basis (see Chapters 8 and 10). Instead, it posts quarterly reports of "Potential Signals of Serious Risks/New Safety Information Identified from the FDA Adverse Event Reporting System (FAERS)," which continues to state that inclusion on this list "does not mean that FDA has determined that the drug has the risk" [61].

In 2016, Bailey *et al*. reviewed the different reporting systems in use worldwide for drug regulatory purposes [62]. They identified 11 international and 97 national-level reporting systems. They found "substantial variability" in the data fields reported, limiting data comparability between systems. However, they were concerned that existing forms focused more on regulatory needs (e.g., drug lot numbers) and recommended developing a form that would

better support clinical care needs. Notably, they identified which countries requested data on such items as rechallenge and dechallenge to allow causality assessment, and called for consensus on data elements used.

Publishers of Reports of Adverse Reactions

The medical literature containing case reports of suspected adverse reactions has largely avoided the issue of causality, although there are many published series of cases reports that apply the Naranjo scoring method [13]. In fact, the *Annals of Pharmacotherapy* now requires that this method (or another validated and appropriate scale) be applied and reported in all case reports published. The majority of single case reports, letters to the editor, or short publications do not provide an explicit judgment using any of the published algorithms. More importantly, despite several meetings and publications, many published reports do not provide information on confounding drug therapy or medical conditions – data elements considered essential for considering causality by experts. This issue was recognized as one of several problems relating to the publication of adverse reactions in the literature and was discussed extensively in 1981 and 1983 at the conferences in Morges, Switzerland. Editors of several medical publications were present and discussed the quality of information in reported cases. The participants developed a list of the types of information desirable for published reports, including data permitting the reader to assess independently the likelihood of the association [63,64]. They concluded that publication of case reports should require the five elements of causality: timing, the nature of the reaction, discontinuation, reintroduction, and alternate causes based on prior history.

The need for publication requirements was underscored in 1990 when Haramburu and her colleagues compared the value of 500 published reports with 500 spontaneous reports with respect to the availability of information needed in most standard causality assessments [65]. Although analysis suggested that the published reports contained significantly more information, the tabulation of these reports indicated very sparse data on both alternate causes/other diseases and other drugs in *both* types of reports. Nonetheless, even years later, few journals appear to require specific types of information for publication of spontaneous reports. This prompted another formal effort in 2004 by the International Society of Pharmacoepidemiology to address this issue. An international working group, looking at the broader need for higher quality publications of suspected adverse reactions to biopharmaceuticals, published recommendations for publications of suspected adverse reactions that were published simultaneously in two major journals focusing on drug safety, *Pharmacoepidemiology & Drug Safety* and *Drug Safety* [66–68]. Subsequently, the recommendations have been adopted by some other journals, including *Annals of Pharmacotherapy* and *Thérapie*.

Methodologic Problems to be Addressed by Pharmacoepidemiologic Research

The goal is to find one or more methods that are reliable, consistent, accurate, and useful for assessing the likelihood of association between an adverse event and a medical product. This problem is compounded by the nature of biopharmaceutical-associated adverse events. They vary in frequency, manifestation, timing relative to exposure, and mechanism. They mimic almost the entire range of human pathology, as well as adding unique new pathologies (e.g., kidney stones consisting of sulindac drug crystals and the oculomucocutaneous syndrome caused by practolol). In addition, biopharmaceutical-associated events are always nested within other pathologies associated with the indication for

the drug. Since drugs are used to produce a beneficial effect, known or expected adverse events are sometimes reluctantly accepted within the clinical risk/benefit equation. However, unknown or unexpected events are inconsistently recognized and/or described. Seldom are the desired baseline and other detailed laboratory measurements or past medical history obtained.

The nature of this task, and its context, has generated two divergent philosophies. One discounts the value or importance of causality assessment of individual reactions, deferring judgment to the results of formal epidemiologic studies or clinical trials [69]. The other contends that evaluation of single reports can help determine at least some degree of association – useful or even critical information when considering ending drug development or a clinical trial or a drug withdrawal [70]. The latter view spurred the evolution of causal evaluation from expert global introspection to structured algorithms and to elaborate probabilistic approaches, as described previously. Further, because of the nature of drug and biologic-associated effects, particularly rare and serious effects, the question has been raised whether epidemiologists need to consider using algorithms or probabilistic methods for case evaluations in formal studies and in clinical trials, since the small numbers available may not be amenable to standard statistical analysis [71].

Currently Available Solutions

There are now a variety of methods for causality assessment of spontaneous reports. Four basic types will be described, chosen as illustrative examples and because they have been widely described in various publications. Agbabiaka and colleagues in a 2008 review concluded that "there is still no method universally accepted for causality assessment of ADRs" [72], as pointed out much earlier by Koch-Weser *et al.* [73]. Khan *et al.* agreed in 2016 [74].

Unstructured Clinical Judgment/Global Introspection

Probably the most common approach to causality assessment is unstructured clinical judgment. An expert is asked to review the clinical information available and to make a judgment as to the likelihood that the adverse event resulted from drug exposure. However, it has been amply demonstrated that global introspection does not work well, for several reasons [7,74].

First, cognitive psychologists have shown that the ability of the human brain to make unaided assessments of uncertainty in complicated situations is poor, especially when assessing the probability of cause and effect, precisely the task of causality assessment [72]. This has been clearly demonstrated for the evaluation of suspected adverse reactions. Several studies have used "expert" clinical pharmacologists to review suspected reactions. Comparing their individual evaluations, these studies documented the extent of their disagreement and illustrated thereby how unreliable global introspection is as a causality assessment method [14–16,69,70,72,73].

Second, global introspection is uncalibrated. One assessor's "possible" might mean the same as another assessor's "probable." This has been well demonstrated in a study of one pharmaceutical company's spontaneous report reviewers, who used both verbal and numerical scales [19]. These and other shortcomings of global introspection as a causality assessment method for adverse reactions are discussed in detail by Lane, Hutchinson, and Kramer, among many others [7,27,72,75–79].

Despite these concerns, global introspection for evaluation of adverse events continues to be used. Most notably, the Uppsala Sweden WHO Centre for Drug Monitoring, which collects the spontaneous reports from national centers

worldwide, has published causality criteria ranging from "certain" to "unassessable/unclassifiable" that essentially represent six levels of global introspection, though they generally incorporate consideration of the more standard criteria for causality [80]. The Portuguese central pharmacovigilance unit (Nucleo de Farmacovigilancia do Centro) utilizes this WHO global introspection (GI) method, in part based upon a comparison of results from evaluation of 200 cases by algorithm methods and the WHO GI method. They found a relatively moderate to high degree of correspondence of judgments for the reactions more likely associated [81].

Algorithm/Criterial Method with Verbal Judgments

The subsequent attempts to address the limitations of global introspection have resulted in the proliferation of methodologic approaches [72,73]. Methods range from simple flowcharts posing 10 or fewer questions to lengthy questionnaires containing up to 84 items. They share a common basic structure essentially based on the original work by Karch and Lasagna [10] and Irey [8,9]– the timing of the adverse event in relation to administration of the drug, alternative etiological candidates, previous recognition of the event as a possible adverse reaction to the drug, the response when dechallenged, and rechallenged. Information relevant to each factor is elicited by a series of questions, the answers to which are restricted to "yes/no" (and for some methods, "don't know").

The advantage of algorithms compared to global introspection [70,76] is improvement in the consistency of ratings among reviewers. Since the consideration of each case is segmented into its components (e.g., timing, confounding diseases, etc.), it allows for a better understanding of areas of disagreement. However, global introspection is still required on the separate elements of the algorithms or decision tables. In some cases, "yes or no" answers are required where a more quantitative estimate of uncertainty would be more appropriate. For example, the reviewer might have to consider whether the appearance of jaundice within one week represented a sufficient duration of drug exposure to be consistent with a drug–event association. Even adherents of some of the methods agree that their procedures for converting answers into probability ratings are arbitrary.

Previously, the FDA used an algorithm based on the Irey and Karch and Lasagna concepts [16]. It inquired sequentially about temporal sequence, dechallenge, rechallenge, and concomitant diseases which might have caused the event. It was tailored for rapid use by professionals with varied backgrounds for the administrative purpose of finding well-documented cases for regulatory signal evaluation. It was also considered useful and easily remembered by clinicians in initial differential diagnosis of a clinical event. However, this very simple approach is less useful for irreversible drug effects (e.g., death, birth defects), since dechallenge and rechallenge are impossible. To address this, an alternate algorithm for fatal outcome events was developed by Turner in the aftermath of the FDA algorithm [17]. Algorithms/decision trees are used by some drug regulatory agencies, such as that of Australia [56].

Algorithms Requiring Scoring of Individual Judgments

Many algorithms permit quantitative judgments by requiring scoring. The answers to the algorithms' questions are converted into a score for each factor, the factor scores are summed, and this overall score is converted into a value on a quantitative probability scale. These

judgments range from the extensive, multiple question method of Venulet [18], which has now been translated for computer use, to the relatively simpler French method [14,60]. The method developed by Kramer *et al*. [12] received considerable review and is representative of the scored methods. Although it was presented in algorithm format with multiple steps, it can also be represented in tabular format.

One of the more practical methods of this type was developed by Naranjo *et al*. [13]. This method has been adopted in a number of clinical settings and by at least one publisher (*Annals of Pharmacotherapy*) and is shown in Figure 29.2 [13]. The RUCAM method was developed in 1994 by Bénichou and Danan [36]. Like the Naranjo method, this has six criteria with three or four levels of scoring for each criterion to derive an overall score. RUCAM has been applied in evaluation of adverse events in HIV clinical trials [82].

CAUSALITY ASSESSMENT
NARANJO SCORED ALGORITHM

QUESTION	ANSWER			SCORE
	Yes	No	Unknown	
Previous reports?	+1	0	0	_____
Event after drug?	+2	−1	0	_____
Event abates on drug removal?	+1	0	0	_____
+ Rechallenge?	+2	−1	0	_____
Alternative causes?	−1	+2	0	_____
Reaction with placebo?	−1	+1	0	_____
Drug blood level toxic?	+1	0	0	_____
Reaction dose-related?	+1	0	0	_____
Past history of similar event?	+1	0	0	_____
ADR confirmed objectively?	+1	0	0	_____
Total Score				

Figure 29.2 A critical scored algorithm illustrated by the method of Naranjo et al. in wide use. This particular method uses some of the basic data elements as well as more details of the history and characteristics of the case, and a score is designated for the response to each question. Source: Reproduced from Naranjo *et al*. [13] with permission of John Wiley & Sons.

These quantitative methods have found applications in a number of settings, ranging from evaluations of suspected adverse reactions by hospital committees (US hospitals are now required by the Joint Commission on Accreditation of Health Care Organizations (JCAHO) to have programs of adverse reaction surveillance) to use by some regulatory authorities, as in France. They are also used, although sometimes only in a research context, by some pharmaceutical manufacturers [21,49]. The specific manner in which they are used has not been well described in the literature.

A newer tool is the Liverpool Adverse Drug Reaction Causality Assessment Tool [83]. The Liverpool tool is a flowchart developed by a team of seven researchers involved in the Adverse Drug Reactions in Children project [84]. They assessed 40 pediatric suspected ADRs with their own methodology (ADRIC Study 1 [85]) and using Naranjo. They then compared the results obtained and carefully examined cases for which assessments differed by more than one degree. This comparison led them to develop a modified version of Naranjo, called the Liverpool ADR Causality Assessment Tool (LCAT), in 2011.

The authors believe that this method may be somewhat cumbersome for practitioners [86]. In 2015, the developers adapted a mobile app for the LCAT and its easy availability may make the Liverpool tool a widely used causality assessment tool [37].

Probabilistic Methods

Recognition of the various methodologic challenges set the stage for the development of an alternative approach: The Bayesian probability approach to assessment of causality (Figure 29.3). This method provided an opportunity for a different perspective on causality assessment, but its difficulty (due to the requirement to use all available information) raised

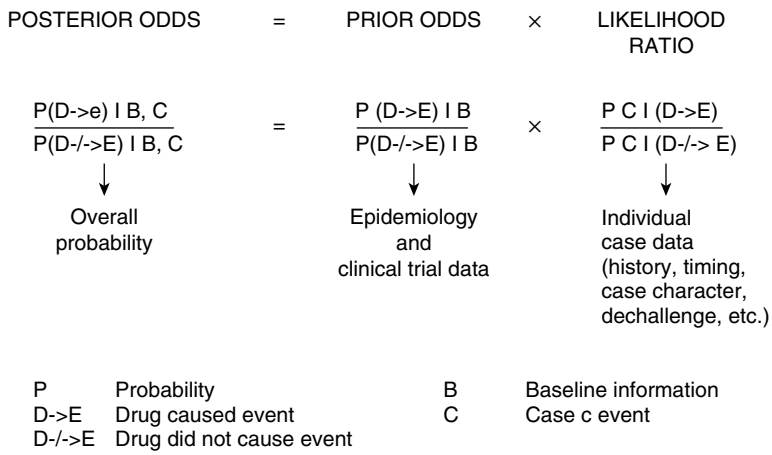


Figure 29.3 The basic equations for the Bayesian analysis of suspected drug-associated events. These provide a structured yet flexible and explicit approach to estimating the probability that an event is associated with one, or more, drugs, as described in the text and extensive literature dating from Auriche [52], Lane *et al*. [28] and others. Since the prior probability estimate is dependent on explicit data from clinical trials and epidemiologic studies, this approach can provide a framework for specific event-related questions in these studies [29].

some new issues about causality assessment of adverse reactions. It also brought adverse reaction evaluation into a larger discussion of the value of applying Bayesian and probabilistic approaches to the analysis of medical and scientific data [3].

First published as a method for adverse reaction assessment by Auriche [52], who participated with Lane and others in a working group within the APWI organization, the Bayesian method was first presented in extensive form in a workshop in 1985 [28]. Several examples were published in a monograph and subsequently in early papers [31,32]. The methods were incorporated into automated versions by both Naranjo and Hutchinson, the latter developing a model using an expert system [87,88]. Naranjo and colleagues implemented a practical spreadsheet/automated version called BARDI (Bayesian Adverse Reaction Diagnostic Instrument) and applied it to a number of practical adverse event problems [70,87,89].

The Bayesian method assesses the probability of an event occurring in the presence of a drug, relative to the probability of that event occurring in the absence of the drug, as illustrated in Figure 29.3. Estimation of this overall probability, the "posterior probability," is based on two components.

- What is known prior to the event, the "prior probability" which is based on clinical trial and epidemiologic data.
- What the likelihoods are, or are not, for the drug to cause the components of the specific case, including its history, timing, characteristics, dechallenge and its timing components, rechallenge, and any other relevant factors.

Application of the Bayesian method requires knowledge of the clinical event, its epidemiology, and relatively specific information about the event's characteristics and kinetics over time. Examples have been published for several types of events, including Stevens–Johnson syndrome, renal toxicity, lithium dermatitis, ampicillin-associated colitis, agranulocytosis, Guillain-Barré syndrome, and pancreatitis [29,78,89,90].

Thus far, this approach appears to be useful for analysis of the perplexing first events in new drug clinical trials, serious spontaneous adverse reaction reports, and possibly rare events discovered in both case–control and cohort pharmacoepidemiologic studies, when standard methods of statistical analysis will not provide sufficient clues as to causality because of inadequate sample size.

Due to automation, the Bayesian method can now be performed rapidly, but the major impediment to is more general application is the frequent lack of information required for robust analyses of events. There are often limited data on the incidence of most events and their occurrence in the presence and absence of most drugs (the required information for the prior probability). There are even fewer data available on the historical risk factors, time course, and specific characteristics of the drug-associated conditions, as opposed to the naturally occurring conditions. However, with the current proliferation of epidemiologic studies, particularly in the areas of natural history of disease as well as of drug-associated diseases such as Stevens–Johnson syndrome, this information is becoming more readily available. So, although this lack of information is sometimes a limitation, it represents both an important challenge and a framework for structuring further understanding.

Bénichou and collaborators have delved further into a mapping process of reactions by type in an attempt to begin classifications of specific drug-associated disease, using acute liver disease as one model that incorporates qualitative clinical definitions of the disease into the judgment [91,92].

For this reason, and with the increasing availability of more epidemiologic data, there are several advantages of using the Bayesian method for analysis of suspected drug-associated events.

- All judgments must be explicit and quantified, permitting better explanations of the degree of uncertainty about each component of information. Further, this approach makes maximum use of the available information and follows the basic rule of not discarding information.
- Since each component is analyzed separately, a sensitivity analysis of each information component can estimate its overall contribution to the final posterior odds or probability estimate. This, in turn, can be used to determine which information is pivotal. For example, if a 10-fold difference in the estimate of the timing does not materially modify the overall posterior odds estimate, further efforts to determine the "best" estimate would not be worthwhile.
- Because of the multistep approach to a judgment, combined with lack of the prejudged weighting present in most other methods, this approach resists the tendency to achieve a result expected on an *a priori* global judgment. This is quite important in evaluating events with multiple possible causes.
- This approach can provide an extensive summary of the information needed and areas needing further research and data compilation.

Thus, the Bayesian approach ultimately provides a "map" to define the information most critical for understanding drug-induced disease and will help formulate the most critical questions to be researched. As disease natural histories and drug-induced diseases are now being described in large population databases, it is essential to link these two types of analyses.

Another useful application of the Bayesian method, combined with the Poisson method for estimating the probability of rare events in populations [33] developed by Bégaud and colleagues, was described by Zapater *et al.* [93]. They demonstrated the feasibility of utilizing both clinical trial and population data to estimate the posterior probabilities of association in complex cases of ticlopidine-associated hepatitis.

More recently, a method based upon the logistic model and directly providing a probability for drug causation ranging between 0 and 1 has been proposed [94]. This method, also proposed in an automated version, preserves some of the basic principles of conditional probabilities; it notably provides a neutral assessment of 0.5 when no information is usable, while conserving the ease of use of classic algorithms. The assessments of seven criteria are combined, the relative weight of each criterion having been calibrated by taking consensual judgment of five experts on a random sample of case reports as gold standard. An improved version was published in 2012 [95].

Comparison Among the Different Methods for Causality Assessment

Several efforts have been made to evaluate and compare methods. The 1983 conference in Crystal City involved the application of several of them to a standardized case, illustrating a considerable lack of concordance for some methods [23].

A much more elegant and detailed evaluation of six representative algorithmic methods was carried out in 1986 by Péré *et al*. [96], who identified standard evaluation criteria and carried out an evaluation of 1134 adverse reactions using the various methods. Significantly, they found only moderate agreement between all pairs, and considerable disagreements on weightings of three of the major criteria – timing, dechallenge, and alternate etiologies – which tends to underline the lack of complete information on the events and their characteristics. More recent attempts to quantify agreements on different methods, including global introspection, have been published by Kramer and Macedo *et al.* [77–79,81].

Given the current state of affairs, where a number of published methods exist, the choice of a method for use in evaluating individual adverse effects will likely be determined by a number of practical factors.

- *How the evaluation will be used*. This refers to both its short-term use (e.g., a rating suggesting more than possible association may be needed to result in a "signal") and long-term use (e.g., will a single highly probable case in a file, not otherwise acted upon, be a source of liability for the evaluator?).
- *The importance of the accuracy of the judgment.* If this evaluation will affect a specific clinical outcome, the continuation of a clinical trial, or the continued marketing of a drug, the accuracy of the judgment may be critical. Conversely, if little hinges upon the judgment, cruder estimates and methods, recognized as such, may suffice.
- *The number of causality evaluations to be made.* The decision on method to use must also be weighed against the time required to make judgments – a concern especially for regulatory agencies and manufacturers where the need for accurate judgments is pitted against the volume of reports. One approach is suggested by the FDA's method of identifying high-priority problems according to their newness and seriousness (see Chapter 8).
- *The accrued value of thorough evaluations*. In some circumstances, the careful, rigorous evaluation of certain categories of drug-associated events will facilitate the more accurate evaluation of subsequent, related events. For example, if a drug under development is anticipated to cause hepatic events, detailed evaluations of hepatic events induced by other drugs may allow more satisfactory causality evaluation of reports received on the new drug [91]. In some cases this results from data collection being focused to a much greater degree, as was initiated in France by Bénichou *et al*., where special reporting forms

using disease-specific criteria for events were developed [91,92]. Zapater *et al.* demonstrated the advantages of these forms in assessing ticlopidine-associated hepatic toxicity, where the evaluation and sensitivity analysis not only clarified the estimated probabilities for the cases, but also suggested that more careful examinations of relative values of hepatic enzymes might further understanding of drug-associated hepatotoxicity [93].

- *Who will perform the evaluation?* Although no specific studies have been carried out to evaluate the interrater differences among differently trained professionals, it is likely that the body of information held by each reviewer will have considerable impact on any of the methods used, including the Bayesian method.

The Future

The field of adverse reaction causality assessment has many unresolved issues, both methodologic and practical, described in the preceding sections. Early on, there was hope for a consensus assessment method [21]. Despite repeated published expressions of need, no standard method has emerged. Several reasons can be suggested.

First, some individuals and institutions have committed to one or a few methods, often through choice of data collecting systems or software [18]. Second, *usability* appears to play a very real role in that choice. Although the Bayesian method was welcomed as the possible gold standard for adverse reaction causality, it was not embraced. Early on, it was difficult to use without automation but even since that barrier was lifted, the Bayesian is often too complex and time-consuming for practical application. The complex Kramer *et al*. algorithm [12] likewise discourages its use in some sectors in spite of improvements from automation, although this has not been documented. Third, concern about possible disagreement with global introspection determinations or algorithm scores has generated concern about potential legal liability [49], since there is no gold standard method.

All these factors suggest the need for considerable further work in several areas.

- Identify the purpose of the causality assessment to better determine the desired rigor, accuracy, and usability of the methods. There will probably always be needs for simpler and rougher methods when large quantities of reports are involved, as well as more complete and rigorous methods when the findings impact the regulatory future of the medical product or its development.
- Identify the critical elements needed for the evaluation of causality for different types of adverse reactions (e.g., hepatic, hematologic, skin etc.) so that this information may be collected at the time of reporting or publishing a spontaneous event. This need has long been recognized [14,23,96] and is being implemented in some centers (e.g., Bordeaux, France; University of Toronto; as well as many pharmaceutical companies) that collect adverse events. Further work in this area can have a major impact on the:
  - collection of better information on the different drug-associated events, using data collection instruments tailored to the event of interest
  - better definition of the dynamics and, ultimately, the pathophysiology and mechanisms of certain types of drug-induced conditions.

At present, with pursuit of the epidemiology and pathophysiology of drug-associated diseases by both individual centers (e.g., the efforts in drug-associated hepatic disease, including liver failure by Lee [97] and the US NIH) and the regulatory agencies, in particular the FDA (pursuing hepatic injury and

other drug-associated disorders such as Stevens–Johnson syndrome), it is likely that such research will support development of much more event-specific methods, such as the ALDEN method for Stevens–Johnson syndrome developed by the European SCAR registry [98].

- Gathering of data on critical elements of the specific adverse events in the course of both clinical trials and epidemiologic studies. Risk factor, history, timing, characteristics, and resolution patterns of adverse events should be described in these studies and incorporated into general data resources on the characteristics of medical events and diseases.
- Further work on automation of the causality evaluation process. Global introspection is still widely used because of the cumbersome nature of many of the more complete methods. Fortunately, several methods are now automated, including the French [96], the Venulet (J. Venulet, personal communication), the Bayesian BARDI [87], the Liverpool tool [20], and the logistic method [94]. Convenient access to the proper questions, arrayed in logical order, as well as background data on the state of information to date, has the potential for radically changing the state of adverse reaction causality evaluation.
- Consideration of new and different methods for assessment. Although future work will likely flow from currently available methods, other approaches have emerged. For example, as part of work on patient safety in the US (see also Chapter 41), "root cause analysis" has emerged to identify the important contributors to adverse events in clinical settings. It examines functional maps of possible contributing factors to identify a cause and determine methods of preventing it. Spath provided one illustration of this approach [99]. Another less generalizable approach described by investigators at the University of Toronto is the N-of-1 trial that evaluates the causality of adverse events in individuals, particularly those with multiple reactions to drugs [100]. Use of an assessment method that can lend itself to modern technology (mobile applications), such as the Liverpool tool, may supersede others simply because of its ease of access [20,37].

In conclusion, how best to assess causality of adverse reactions continues to represent a challenge. With increased understanding that causality assessment is necessary in the regulatory and drug development and testing processes, the need for consensus on one or more methods, depending on use, continues. Application of detailed causality assessment, particularly when viewed prospectively with collection of data in both pharmacovigilance centers and clinical studies, can ultimately contribute to the overall understanding of many drug-associated diseases.

References

1 Jones JK. Uses of drug–event assessment. *Drug Inf J* 1984; **18**: 233–40.

2 Yerushalmy J, Palmer CE. On the methodology of investigations of etiologic factors in chronic diseases. *J Chronic Dis* 1959; **10**(1): 27–40.

3 Feinstein AR. Clinical Biostatistics. XLVII. Scientific standards vs. statistical associations and biologic logic in the analysis of causation. *Clin Pharmacol Ther* 1979; **25**: 481–92.

4 Anonymous. Causality assessment of an adverse event following immunization. World Health Organization. www.who.int/vaccine_safety/publications/gvs_aefi/en/ (accessed April 19, 2019).

5 Shakir SAW, Layton D. Causal association in pharmacovigilance and pharmacoepidemiology. *Drug Saf* 2002; **25**: 467–71.

6 Tubert P, Bégaud B, Haramburu F, *et al*. Spontaneous reporting: how many cases are

required to trigger a warning? *Br J Clin Pharmacol* 1991; **32**: 407–8.

7 Hutchinson TA, Lane DA. Assessing methods for causality assessment. *J Clin Epidemiol* 1989; **42**: 5–16.

8 Irey NS. Diagnostic problems in drug-induced diseases. In: Meyler L, Peck HM, eds. *Drug-Induced Diseases*, vol. IV. Amsterdam: Elsevier Science Publishers, 1972, pp. 1–24.

9 Irey NS. Adverse drug reactions and death: a review of 827 cases. *JAMA* 1976; **236**: 575–8.

10 Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977; **21**: 247–54.

11 Aronson JK, Hauben M. Anecdotes that provide definitive evidence. *BMJ* 2006; **333**(7581): 1267–9.

12 Kramer MS, Leventhal JM, Hutchinson TA, Feinstein AR. An algorithm for the operational assessment of adverse drug reactions. I. Background, description, and instructions for use. *JAMA* 1979; **242**: 623–32.

13 Naranjo CA, Busto U, Sellers EM, *et al*. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–45.

14 Bégaud B, Evreux JC, Jouglard J, Lagier G Unexpected or toxic drug reaction assessment (Imputation). Actualization of the method used in France. Thérapie 1985; **40**: 111–18.

15 Emanueli A, Sacchetti G. An algorithm for the classification of untoward events in large scale clinical trials. *Agents Actions* 1980; **7**: 318–22.

16 Jones JK. Adverse reactions in the community health setting: approaches to recognizing, counseling, and reporting. *Fam Commun Health* 1985; **5**: 58–67.

17 Turner WM. The Food and Drug Administration algorithm. *Drug Inf J* 1984; **18**: 259–66.

18 Venulet J, Ciucci AG, Berneker GC. Updating of a method for causality assessment of adverse drug reactions. *Int J Clin Pharmacol Ther Toxicol* 1986; **24**: 559–68.

19 Castle WM. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 297–302.

20 Gallagher RM, Kirkham JJ, Mason JR, *et al*. Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. *PLoS One* 2011; **6**(12): e28096.

21 Venulet J, Berneker GC, Ciucci AG, eds. Assessing Causes of Adverse Drug Reactions, London: Academic Press, 1982.

22 Herman RL, ed. Drug–event associations: perspectives, methods, and users. Proceedings of Drug Information Association Workshop, Arlington, VA, 1983.

23 Jones JK. Drug–event associations: a view of the current status. Epilogue. *Drug Inf J* 1984; **18**: 233–40.

24 Jones JK, Herman RL. Assessing adverse drug experiences (letter). *Ann Intern Med* 1988; **108**: 160.

25 Venulet, J. APWI: Past & Present. *Drug Inf J* 1991; **25**: 229–33.

26 Herman RL, Fourrier A. *List of References: Third International APWI Meeting*. Paris, November 27–28, 1992.

27 Lane DA. A probabilist's view of causality assessment. *Drug Inf J* 1984; **18**: 323–30.

28 de Finetti B. *Theory of Probability*. New York: John Wiley, 1974.

29 Jones JK, Herman RL (guest eds). Introduction, the future of adverse drug reaction diagnosis: computers, clinical judgement and the logic of uncertainty. Proceedings of the Drug Information Association Workshop, Arlington, Virginia, February 1986. *Drug Inf J* 1986; **20**: 383.

30 Lane D. Causality assessment for adverse drug reactions: a probabilistic approach. In: Berry D, ed. *Statistical Methodology in the Pharmaceutical Sciences*. New York: Marcel Dekker, 1990, pp. 475–507.

31 Lane DA, Kramer MS, Hutchinson TA, Naranjo C, Jones JK, The causality assessment of adverse drug reactions using a Bayesian approach. *Pharm Med* 1987; **2**: 265–83.

32 Jones, JK. A Bayesian approach to causality assessment. *Psychopharm Bull* 1987; **23**: 395–9.

33 Bégaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? *Br J Clin Pharmacol* 1994; **38**: 401–4.

34 Bénichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs – I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J Clin Epidemiol* 1993; **46**: 1323–30.

35 Bénichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs – II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. *J Clin Epidemiol* 1993; **46**: 1331–6.

36 Bénichou C, Danan G. A new method for drug causality assessment: RUCAM. In: *Adverse Drug Reactions. A Practical Guide to Diagnosis and Management*. New York. John Wiley & Sons Ltd, 1994, pp. 277–84.

37 Ithnin M, Mohd Rani MD, Abd Latif Z, *et al*. Mobile app design, development, and publication for adverse drug reaction assessments of causality, severity, and preventability. *JMIR Mhealth Uhealth* 2017; **5**(5): e78.

38 Moore N, Paux G, Bégaud B, *et al*. Adverse drug reaction monitoring: doing it the French way. *Lancet.* 1985; **2**:1056–1058.

39 CHMP Working Party on Pharmacovigilance. *Procedure for Causality Classification in Pharmacovigilance in the European Community*. EC Document III/3445/91–EN, July 1991.

40 Miremont G, Haramburu F, Bégaud B, Péré JC, Dangoumau J. Adverse drug reactions: physicians' opinions versus a causality assessment method. *Eur J Clin Pharmacol* 1994; **46**: 285–9.

41 Mascolo A, Scavone C, Sessa M, *et al*. Can causality assessment fulfill the new European definition of adverse reaction? A review of methods used in spontaneous reporting. *Pharmacol Res* 2017; **123**: 122–9.

42 Jones, JK. Causality assessment of suspected adverse drug reactions: a transatlantic view. *Pharmacoepidemiol Drug Saf* 1992; **1**: 251–60.

43 Pless RP. Vaccine-associated adverse events: surveillance and reporting. In: Krogh C, ed. *Compendium of Pharmaceuticals and Specialities*, 33rd edn. Ottawa: Canadian Pharmaceutical Association, 1998.

44 Collet JP, MacDonald N, Cashman N, Pless R, for the Advisory Committee on Causality Assessment. Monitoring signals for vaccine safety: the assessment of individual adverse event reports by an expert advisory committee. *Bull WHO* 2000; **78**(2): 178–85.

45 US Code of Federal Regulations 1987; 21: 312.22 on Investigational New Drug Procedures.

46 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Guidance for Industry and Investigators, *Safety Reporting Requirements for INDs and BA/BE Studies*. December 2012.

47 US FDA, Final Rule. *Investigational New Drug Safety Reporting Requirements for Human Drug and Biological Products and Safety Reporting Requirements for Bioavailability and Bioequivalence Studies in Humans*, 75 FR 59935 (September 29, 2010). Before 1985 US Code of Federal Regulations (CFR) 21 CFR §310.300; and currently, 21 CFR §314.80.

48 US FDA, Final Rule. *Requirements on Content and Format of Labeling for Human Prescription Drug and Biological Products*, 71 FR 3922–3997 (January 24, 2006).

49 Castle W, Baker A. Communication of ADR information: the viewpoint of a British company. *Drug Inf J* 1985; **19**: 375–80.

50 Stephens M. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 307–14.

51 Stephens M. *The Detection of New Adverse Drug Reactions*. London: Plenum Press, 1985.

52 Auriche M. Approache bayésienne de l'imputabilité des phénomènes indésirables aux médicaments. *Thérapie* 1985; **40**: 301–6.

53 Freilich WB. Legal perspectives in causality assessment. *Drug Inf J* 1984; **18**: 211–17.

54 Margulies JB. Epidemiologic causation in the courtroom: square pegs in round holes? *Drug Inf J* 1991; **25**: 217–27.

55 Henderson, TW. Legal aspects of disease clusters. Toxic tort litigation: medical and scientific principles in causation. *Am J Epidemiol* 1990; **132**: S69–78.

56 Mashford ML. The Australian method of drug–event assessment. *Drug Inf J* 1984; **18**: 271–3.

57 Wiholm BE. The Swedish drug–event assessment methods. *Drug Inf J* 1984; **18**: 267–9.

58 Dangoumau J, Bégaud B, Boisseau A. Les effets indesirables des medicaments. *Presse Med* 1980; **9**:1607–9.

59 Miremont-Salamé G, Théophile H, Haramburu F, Bégaud B. Causality assessment in pharmacovigilance: the French method and its successive updates. *Thérapie* 2016; **71**: 179–86.

60 Théophile H, Dutertre JP, Gérardin M, *et al*. Validation and reproducibility of the updated French causality assessment method: an evaluation by pharmacovigilance centres and pharmaceutical companies. *Thérapie* 2015; **70**(5): 465–76.

61 Potential Signals of Serious Risks/New Safety Information Identified from the FDA Adverse Event Reporting System (FAERS), www.fda.gov/Drugs/GuidanceComplianceRegulatory Information/Surveillance/AdverseDrugEffects/ucm082196.htm (accessed April 19, 2019).

62 Bailey C, Peddie D, Wickham ME, *et al*. Adverse drug event reporting systems: a systematic review. *Br J Clin Pharmacol* 2016; **82**(1): 17–29.

63 Editorial. Improving reports of adverse drug reactions. *BMJ* 1984; **289**: 898–9.

64 Venulet J. Incomplete information as a limiting factor in causality assessment of adverse drug reactions and its practical consequences. *Drug Inf J* 1986; **20**: 423–31.

65 Haramburu F, Bégaud, B, Péré JC. Comparison of 500 spontaneous and 500 published reports of adverse reactions. *Eur J Clin Pharmacol* 1990; **39**: 287–8.

66 Kelly WN, Arellano FM, Barnes J, *et al*. Guidelines for submitting adverse event reports for publication. *Drug Saf* 2007; **30**: 367–73.

67 Kelly WN, Arellano FM, Barnes J, *et al*. Guidelines for submitting adverse event reports for publication. *Pharmacoepidemiol Drug Saf* 2007; **16**: 581–7.

68 Kelly W, Arellano F, Barnes J, et al. Guidelines for submitting adverse event reports for publication.*Thérapie* 2009; **64**(4): 289–94.

69 Louik C, Lacouture PG, Mitchell AA, *et al*. A study of adverse reaction algorithms in a drug surveillance program. *Clin Pharmacol Ther* 1985; **38**: 183–7.

70 Naranjo, CA, Lanctot, KL, Lane DA. The Bayesian differential diagnosis of neutropenia and antiarrhythmic agents. *J Clin Pharmacol* 1990; **30**: 1120–7.

71 Jones JK. Epidemiologic perspective on causality assessment for drug associated events. *Drug Inf J* 1986; **20**: 413–22.

72 Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. *Drug Saf* 2008; **31**(1): 21–37.

73 Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. *Eur J Clin Pharmacol* 1977; **11**: 75–8.

74 Khan LM, Al-Harthi SE, Osman AM, Sattar MA, Ali AS. Dilemmas of the causality assessment tools in the diagnosis of adverse

drug reactions. *Saudi Pharm J* 2016; **24**(4): 485–93.

75 Arimone Y, Bégaud B, Miremont-Salamé G, *et al*. Agreement of expert judgment in causality assessment of adverse drug reactions. *Eur J Clin Pharmacol* 2005; **61**(3): 169–73.

76 Karch FE, Smith CL, Kerzner B, Mazzullo JM, Weintraub M, Lasagna L. Adverse drug reactions – a matter of opinion. *Clin Pharmacol Ther* 1976; **19**: 489–92.

77 Kramer MS. Assessing causality of adverse reactions: global introspection and its limitations. *Drug Inf J* 1986; **20**: 433–7.

78 Macedo AF, Marques FB, Ribeiro CF, Texeira F. Causality assessment of adverse drug reactions: comparisons of the results obtained from published decisional algorithms and from the evaluations of an expert panel, according to different levels of imputability. *J Clin Pharmacol Therapeut* 2003; **28**: 137–43.

79 Macedo AF, Marques FB, Ribeiro CF. Can decisional algorithms replace global introspection in the individual causality assessment of spontaneously reported ADRs? *Drug Saf* 2006; **29**(8): 697–702.

80 Meyboom RH. Causality assessment revisited. *Pharmacoepidemiol Drug Saf* 1998; **7**(Suppl 1): S63–5.

81 Macedo AF, Marques FB, Ribeiro CF, Teixeira F. Causality assessment of adverse drug reactions: comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel. *Pharmacoepidemiol Drug Saf* 2005; **14**(12): 885–90.

82 Bonfanti P, Valsecchi L, Parazzini F, et al. Relazione di causalitá delle reaziioni averse a farmaci: un'esperienza nell'utilizzo degli inibitori della protease di HIV. *Clin Ter* 2000; **151**: 411–15.

83 Gallagher RM, Kirkham JJ, Mason JR, *et al*. Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. *PLoS One* 2011; **6**(12): e28096.

84 Smyth RL, Peak M, Turner MA, *et al*. *ADRIC: Adverse Drug Reactions In Children – a programme of Research Using Mixed Methods*. Southampton: NIHR Journals Library, 2014.

85 Gallagher RM, Bird KA, Mason JR*et al.* Adverse drug reactions causing admission to a paediatric hospital: a pilot study. *J Clin Pharm Therapeut* 2011; **36**: 194–9

86 Mouton JP, Mehta U, Rossiter DP, Maartens G, Cohen K. Interrater agreement of two adverse drug reaction causality assessment methods: a randomised comparison of the Liverpool Adverse Drug Reaction Causality Assessment Tool and the World Health Organization–Uppsala Monitoring Centre system. *PLoS One* 2017; **12**(2): e0172830.

87 Naranjo CA, Lanctot, KL. Microcomputer assisted Bayesian differential diagnosis of severe adverse drug reactions to new drugs: a 4-year experience. *Drug Inf J* 1991; **25**: 243–50.

88 Hutchinson TA, Dawid AP, Spiegelhalter DJ Cowell RG, Roden, S. Computerized aids for probabilistic assessment of drug safety I: a spreadsheet program. *Drug Inf J* 1991; **25**: 29–39.

89 Lanctot KL, Naranjo, CA. Using microcomputers to simplify the Bayesian causality assessment of adverse drug reactions. *Pharm Med* 1990; **4**: 185–95.

90 Hauben M. Application of an empiric Bayesian data mining algorithm to reports of pancreatitis associated with atypical antipsychotics. *Pharmacotherapy* 2004; **24**(9): 1122–9.

91 Bénichou C. *Guide Practique de Pharmacovigilance*. Paris: Editions Pradel, 1992.

92 Bénichou C, Danon G. Experts' opinion in causality assessment: results of consensus meetings. *Drug Inf J* 1991; **25**: 251–5.

93 Zapater P, Such J, Perez-Mateo M, Horga J F. A new Poisson and Bayesian-based method to assign risk and causality in patients with suspected hepatic adverse drug reactions. *Drug Saf* 2002; **25**: 735–50.

94 Arimone Y, Bégaud B, Miremont-Salamé G, *et al*. A new method for assessing drug causation provided agreement with expert judgment. *J Clin Epidemiol* 2008; **59**: 308–14.

95 Théophile H, André M, Arimone Y, Haramburu F, Miremont-Salamé G, Bégaud B. An updated method improved the assessment of adverse drug reactions in routine pharmacovigilance. *J Clin Epidemiol* 2012; **65**: 1069–77.

96 Péré JC, Bégaud B, Haramburu F, Albin H. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986; **40**: 451–61.

97 Lee WM. Drug-induced hepatotoxicity. *N Engl J Med* 2003; **349**: 474–85.

98 Sassolas B, Haddad C, Mockenhaupt M, *et al*. ALDEN, an algorithm for assessment of drug causality in Stevens–Johnson Syndrome and toxic epidermal necrolysis: comparison with case–control analysis. *Clin Pharmacol Ther* 2010; **88**(1): 60–8.

99 Spath P. Uncover root causes with E&CF charts: illustrate causal factors in accident sequence. www.reliasmedia.com/articles/46110-uncover-root-causes-with-e-amp-cf-charts (accessed April 19, 2019).

100 Knowles AR, Uetrecht JP, Shear NH. Confirming false adverse reactions to drugs by performing individualized, randomized trials. *Can J Clin Pharmacol* 2002; **9**: 149–53.

30

Molecular Pharmacoepidemiology

Christine Y. Lu[1] *and Stephen E. Kimmel*[2]

[1] *Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, MA, USA*
[2] *Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*

Precision medicine has been defined by the National Institutes of Health (NIH) in the United States as an "approach to disease prevention and treatment based on people's individual differences in environment, genes and lifestyle." Genomic technologies are increasingly available and their use in clinical care has grown substantially over the last decade. As of July 10, 2018, there are over 54 852 tests for 11 233 conditions, 16 433 genes, and 507 laboratories according to the NIH's Genetic Testing Registry. There are different types and applications of genomic technologies. These include disease screening (in asymptomatic individuals, e.g., newborn screening and carrier screening), diagnosis (in symptomatic individuals, e.g., identifying genetic variants in children with suspected genetic disorders), prognostic (e.g., Lynch syndrome for increased risk for colorectal, endometrial, ovarian, and other cancers associated with mutations in mismatch-repair genes), risk assessment or susceptibility (e.g., *BRCA1/BRCA2* testing for hereditary breast and ovarian cancer), inform reproductive choices (prenatal and preimplantation testing), and pharmacogenetic (predict treatment response or adverse events, e.g., HLAB*1502 for risk of Stevens–Johnson syndrome and toxic epidermal necrolysis and to guide use of carbamazepine).

One of the most challenging aspects of precision medicine that involves pharmaceuticals is to understand why individuals and groups of individuals respond differently to a specific drug therapy, in terms of both beneficial and adverse effects. Reidenberg observes that, while the prescriber has basically two decisions to make while treating patients (i.e., choosing the right drug and choosing the right dose), interpreting the interindividual variability in outcomes of drug therapy includes a much wider spectrum of variables, including the patient's health profile, prognosis, disease severity, quality of drug prescribing and dispensing, adherence with prescribed drug regimen (see Chapter 38), and last but not least, the genetic profile of the patient [1].

The effects of genes and other biomarkers on drug response can be studied using molecular pharmacoepidemiology. This is the study of the manner in which molecular biomarkers alter the clinical effects of medications in populations. Just as the basic science of pharmacoepidemiology is epidemiology, applied to the

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

content area of clinical pharmacology, the basic science of molecular pharmacoepidemiology is epidemiology in general and molecular epidemiology specifically, also applied to the content area of clinical pharmacology. Thus, many of the methods and techniques of epidemiology apply to molecular pharmacoepidemiologic studies. However, there are several features of molecular pharmacoepidemiology that are somewhat unique to the field, as discussed later in this chapter. Most of the discussion will focus on studies related to genes, but the methodologic considerations apply equally to studies of proteins (e.g., proteomics) and other biomarkers, such as the microbiome (the genes within the microbial cells, primarily bacteria in the gut, harbored within each person) and mRNA (messenger RNA that results from DNA transcription).

It has been suggested that, on average for each medication, about one out of three treated patients experiences beneficial effects, one out of three does not show the intended beneficial effects, 10% experience only side effects, and the rest of the patient population is nonadherent so that the response to the drug is difficult to assess [2]. Although this is just a crude estimate, it highlights the challenge of individualizing therapy in order to produce a maximal beneficial response and minimize adverse effects. Although it is clear that many factors can influence medication efficacy and adverse effects, including age, drug interactions, and medication adherence (see Chapter 2), genetics can clearly be an important contributor in the response of an individual to a medication. Genetic variability can account for a large proportion (e.g., some estimates range from 20% to 95% [3]) of variability in drug disposition and medication effects [3–7].

In addition to altering dosing requirements, genetics can influence response to therapy by altering drug targets or the pathophysiology of the disease states that drugs are used to treat [8–13].

Genetic Variability in Drug Response: Historical Perspective

Although molecular pharmacoepidemiology is a relatively new area of research, the idea that different individuals have different susceptibility to the effects of medications is not new. Since the advent of modern drugs soon after the Second World War, physicians, pharmacists, and patients have been confronted with interindividual variability in the effects of drug therapy. Some patients need higher than normal doses to achieve an optimum effect. In other patients, unwanted and adverse effects occur even in low doses, while some patients receive no apparent effect of the medication at all. History shows a number of cases where genetics or factors that may be correlated with genetic variability played a role in interpreting and predicting drug effects (Box 30.1).

One of the best-known "classic" examples of genetic variance in drug response is the metabolic defect caused by glucose-6-phosphate dehydrogenase (G6PD) deficiency [14]. This X-linked chromosome disorder is present in about 10% of African men, and occurs at low expressed frequencies in some Mediterranean peoples. In carriers of this deficiency, hemolytic reaction occurs after exposure to oxidant drugs such as antimalarials (e.g., chloroquine), but is also seen

Box 30.1 Some examples of "old" clinically relevant gene–drug interactions

Hemolysis in patients exposed to antimalarial therapy and G6PD deficiency [14]
Prolonged action of suxamethonium due to plasma cholinesterase polymorphism [15]
Neuropathy in patients exposed to isoniazid N-acetyltransferase polymorphism [16]
Inefficacy of codeine as analgesic in poor metabolizers (CYP2D6) [17]

in patients using drugs such as aspirin, probenecid, or vitamin K.

Another early stimulus for pharmacogenetic thinking was the observation that, in the 1 in 3500 white subjects who are homozygous for the gene encoding an atypical form of butyrylcholinesterase, the inability to sufficiently hydrolyze the muscle relaxant drug succinylcholine could lead to prolonged, drug-induced muscle paralysis resulting in severe, frequently fatal, apnea [15].

A third pharmacogenetic antecedent is the example of drug-induced neuropathy in patients with low activity levels of the metabolic enzyme N-acetyltransferase [16]. This enzyme plays an important role in Phase II pathways of drug metabolism, and variance of the activity of this enzyme may lead to dramatic and clinically relevant differences in the plasma concentrations of drugs such as isoniazid, hydralazine, and procainamide.

A final example is the metabolic variance caused by one of the many cytochrome P450 enzymes (CYP). Doctors treating patients with codeine as an analgesic have observed for decades that some patients do not respond at all to normal doses. These clinical observations were not well understood until it was discovered that a polymorphism of CYP2D6 (a subfamily of cytochrome P450) could result in suboptimal transformation of the inactive prodrug codeine into the active form, morphine [17]. The example of codeine points to inherited lack of efficacy. However, genetic polymorphisms of CYP2D6 also have consequences for drug safety, as discussed later in this chapter.

Definitions and Concepts

Genetic Variability

Building on the success of the various human genome initiatives, it is now estimated that there are approximately 25 000 regions of the human genome that are recognized as genes because they contain deoxyribonucleic acid (DNA) sequence elements including exons (sequences that encode proteins), introns (sequences between exons that do not directly encode amino acids), and regulatory regions (sequences that determine gene expression by regulating the transcription of DNA to RNA, and then the translation of RNA to protein). Some of these sequences have the ability to encode RNA (ribonucleic acid, the encoded messenger of a DNA sequence that mediates protein translation) and proteins (the amino acid sequence produced by the translation of RNA). In addition, we are learning a great deal about genomic regions that do not encode RNA or protein, but play important roles in gene expression and regulation such as epigenetics (changes in DNA expression that occur but are not related to the base order, such as DNA methylation). Moreover, changes in the DNA of microbial cells (the microbiome) can influence human response to medications.

Thanks to numerous human genome initiatives, we also have substantial information about interindividual variability in the human genome. The most common form of genomic variability is a single nucleotide polymorphism (SNP), which represents a substitution of one nucleotide (i.e., the basic building block of DNA, also referred to as a "base") for another, which is present in at least 1% of the population. Each person has inherited two copies of each allele (one from the paternal chromosome and one from the maternal chromosome). The term *allele* refers to the specific nucleotide at one point in the genome inherited from either the father or mother, and the combination of alleles in an individual is called a genotype. When the two alleles are identical (i.e., the same nucleotide sequence on both chromosomes), the genotype is referred to as "homozygous" and when the two alleles are different (i.e., different nucleotide sequences on each chromosome), the genotype is referred to as "heterozygous."

Approximately 10 million SNPs are thought to exist in the human genome, with an estimated two common missense (i.e., amino acid changing) variants per gene [18].

However, SNPs are not the only form of genetic variation that may be relevant to human traits and diseases. For example, copy number variants (CNV), sections of the genome that have repeats of base pairs, may also have a role in disease etiology [19]. DNA methylation, where methyl groups are added to DNA, thus changing the activity of DNA (which itself is regulated by genetics), and variability in the gut microbiome can also alter drug response [20–22].

Finally, we also recognize that the genome is not simply a linear nucleotide sequence, but that population genomic structure exists in which regions as large as 100 kilobases (a kilobase being a thousand nucleotides, or bases) in length-defined units remain intact over evolutionary time [23]. These regions define genomic block structures that may define haplotypes, which are sets of genetic variants that are transmitted as a unit across generations.

Thus, the complexity of genome structure and genetic variability that influences responses to medications provides unique challenges to molecular pharmacoepidemiology.

Pharmacogenetics and Pharmacogenomics

While the term *pharmacogenetics* is predominantly applied to the study of how genetic variability is responsible for differences in patients' responses to drug exposure, the term *pharmacogenomics*, as well as including studies of genetic variability on drug response, also encompasses approaches simultaneously considering data about thousands of genotypes, as well as responses in gene expression to existing medications [24,25]. Although the term *pharmacogenetics* is sometimes used synonymously with pharmacogenomics, the former usually refers to a candidate-gene approach as opposed to a genome-wide approach in pharmacogenomics (both discussed later in this chapter).

The Interface of Pharmacogenetics and Pharmacogenomics with Molecular Pharmacoepidemiology

Pharmacogenetic and pharmacogenomic studies usually are designed to examine intermediate endpoints between drugs and outcomes (such as drug levels, pharmacodynamic properties, or surrogate markers of drug effects) and often rely on detailed measurements of these surrogates in small groups of patients in highly controlled settings. Molecular pharmacoepidemiology focuses on the effects of genetics on clinical outcomes and uses larger observational and experimental methods to evaluate the effectiveness and safety of drug treatment in the population. Molecular pharmacoepidemiology uses similar methods to pharmacoepidemiology to answer questions related to the effects of genes on drug response. Thus, molecular pharmacoepidemiology answers questions related to:

- the population prevalence of SNPs and other genetic variants
- evaluating how these genetic variants alter disease outcomes
- assessing the impact of gene–drug and gene–gene interactions on drug response and disease risk
- evaluating the usefulness and impact of genetic tests in populations exposed, or to be exposed, to drugs (i.e., comparative effectiveness, see Chapter 26).

There are, however, some aspects of molecular pharmacoepidemiology that differ from the rest of pharmacoepidemiology. These include the need to understand the complex relationship between medication response and the vast number of potential molecular and genetic

influences on this response; a focus on interactions among these factors and interactions between genes and environment (including other medications) that raises issues of sample size and has led to interest in novel designs; and the need to parse out the most likely associations between genes and drug response from among the massive number of potentially important genes identified through bioinformatics (the science of developing and utilizing computer databases and algorithms to accelerate and enhance biological research).

As stated previously, the basic science of epidemiology underlies molecular pharmacoepidemiology just as it underlies all pharmacoepidemiology. What is different is the need for approaches that can deal with the vast number of potential genetic influences on outcomes; the possibility that "putative" genes associated with drug response may not be the actual causal genes, but rather a gene near or otherwise associated with the causal gene on the chromosome in the population studied (and that may not be similarly linked in other populations); the potential that multiple genes, each with a relatively small effect, work together to alter drug response; and the focus on complex interactions between and among genes, drugs, and environment. By discussing the potential approaches to these challenges in this chapter, it is hoped that both the similarities and differences between pharmacoepidemiology and molecular pharmacoepidemiology will be made clear.

Clinical Problems to be Addressed by Pharmacoepidemiologic Research

It is useful to conceptualize clinical problems in molecular pharmacoepidemiology by thinking about the mechanism by which genes can affect drug response.

The effect that a medication has on an individual can be affected at many points along the pathway of drug distribution and action. This includes absorption and distribution of medications to the site of action, interaction of the medication with its targets, metabolism of the drug, and drug excretion (see Chapter 2) [5,24–26]. These mechanisms can be categorized into three general routes by which genes can affect a drug response: pharmacokinetic, pharmacodynamic, and gene–drug interactions in the causal pathway of disease. These will be discussed in turn below.

Pharmacokinetic Gene–Drug Interactions

Genes may influence the pharmacokinetics of a drug by altering its metabolism, absorption, or distribution. As discussed previously, the fact that different individuals might metabolize medications differently has been well known for decades (see also Chapter 2). Metabolism of medications can either inactivate their effect or convert an inactive prodrug into a therapeutically active compound. Drugs can be metabolized either through Phase I reactions (oxidation, reduction, and hydrolysis) or Phase II (conjugation) reactions (e.g., methylation) [27]. The genes that are responsible for variable metabolism of medications are those that code for various enzyme systems, especially the cytochrome P450 enzymes.

The gene encoding CYP2D6 represents a good example of the various ways in which polymorphisms can alter drug response. Some of the genetic variants lead to low or no activity of the CYP2D6 enzyme whereas some individuals have multiple copies of the gene, leading to increased metabolism of drugs. A specific example is the clinically relevant association between polymorphism of CYP2D6 and the risk of antipsychotic-induced extrapyramidal syndromes, as measured by the need for antiparkinsonian medication. In a case–control study by Schillevoort *et al.*, patients using the CYP2D6-dependent antipsychotic drugs (e.g., haloperidol) who were poor metabolizers were

more than four times more likely to need antiparkinsonian medication than the extensive metabolizers (odds ratio 4.4; 95% confidence interval [CI] 1.1–17.7) [28]. An increased risk was not observed for patients using non-CYP2D6-dependent antipsychotic drugs (odds ratio 1.2; 95% CI 0.2–6.8). The decreased metabolic activity of CYP2D6 may also lead to lower drug efficacy, as illustrated previously for codeine, which is a prodrug that is metabolized to the active metabolite, morphine, by CYP2D6 [17,29]. It has been estimated that approximately 6–10% of Caucasians have variants that result in CYP2D6 genotypes that encode dysfunctional or inactive CYP2D6 enzyme, in whom codeine is an ineffective analgesic [9].

There is important interethnic variability of CYP2D6 alleles and phenotypes. An analysis of CYP2D6 allele-frequency data from >60 000 individuals suggests that diplotype frequencies predicting poor metabolism are highest among Europeans and the Ashkenazi Jewish population (about 5–6%) and lowest among East and South Central Asians, Oceanians, and Middle Easterns (<1% in each of these populations) [30]. In contrast, diplotype frequencies predicting ultrarapid metabolism were highest in Oceanians (21.2%), followed by Ashkenazi Jews and Middle Easterns (about 11% in each of these populations) and lowest in East Asians (1.4%).

Many drug–CYP2D6 genetic variant interactions have been reported based on experimental or epidemiologic associations. CYP2D6 is one of the most common pharmacogenomic markers included in drug labeling by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA). However, predicting clinical outcomes in daily practice based on CYP2D6 genetic data in a valid fashion remains complex, with probably an exception for optimizing breast cancer treatment with tamoxifen by assessing CYP2D6-metabolizing state before initiating therapy [31–33]. Drug–gene associations shown in one study cannot always be replicated in another [34]. Obviously, variance in drug response has many determinants and singling out only one genetic factor fails to account for the co-occurrence, interplay, and interactions of several other factors (e.g., disease severity, exposure variability over time, physiologic feedback mechanisms, testing bias), that are also of critical importance for molecular pharmacoepidemiology [35].

The genetic polymorphism of thiopurine methyltransferase (TPMT) in treating cancer patients is another example [10,36,37]. In its usual state, TPMT metabolizes thiopurine drugs, which would otherwise be toxic if not excreted. In approximately 86–97% of individuals, TPMT activity is high and allows normal drug excretion. In 3–14% activity is intermediate due to the presence of a heterozygous variant in the TPMT gene. In 0.03–0.56%, activity is so low (due to a homozygous variant in the TPMT gene) that patients using drugs such as azathioprine, mercaptopurine, or thioguanine accumulate excessive concentrations of the active thioguanine nucleotides, leading to severe hematologic toxicity. Thus, TPMT genotyping prior to treatment with these agents can be useful to avoid potential toxicities [38,39]. Alternatively, given the rarity of the homozygous variants, individuals who experience treatment-related toxicities may be genotyped for TPMT, and this may influence the course of further treatments.

In addition to metabolism, genes that alter the absorption and distribution of medications may also alter drug levels at tissue targets. These include, for example, genes that code for transporter proteins such as the ATP-binding cassette transporter proteins (ABCB, also known as the multidrug-resistance [MDR]-1 gene) [40], which has polymorphisms that have been associated with, for example, resistance to antiepileptic drugs [41]. It has been found that patients with drug-resistant epilepsy (approximately one of three patients with epilepsy is a nonresponder) are more likely to have the CC polymorphism of ABCB1, which is associated with increased

expression of this transporter drug-efflux protein (odds ratio 2.66; 95% CI 1.32–5.38) [41]. Of note, and consistent with the complexities of molecular pharmacoepidemiologic research noted later, the ABCB1 polymorphism falls within an extensive block of linkage disequilibrium (LD). LD is defined by a region in which multiple genetic variants (e.g., SNPs) are correlated with one another due to population and evolutionary genetic history. As a result, an SNP may be statistically associated with disease risk, but is also in LD with the true causative SNP. Therefore, the SNP under study may not itself be causal but simply linked to a true causal variant [41]. One of the major challenges in genetics research at this time is developing methods that can identify the true causal variant(s) that may reside in an LD block.

Pharmacodynamic Gene–Drug Interactions

Once a drug is absorbed and transported to its target site, its effect may be altered by differences in the response of drug targets. Therefore, polymorphisms in genes that code for drug targets may alter the response of an individual to a medication.

For example, polymorphisms of the beta-2-adrenergic receptor (beta-2-AR) might affect response to beta-agonists (e.g., albuterol) in asthma patients. In particular, the coding variants at position 16 within the beta-2-AR gene (beta-2-AR-16) have been suggested to determine patient response to albuterol treatment [11]. Israel *et al*. showed that the Arg-Arg genotype at beta-2-AR-16 was positively associated with clinical response to albuterol in patients who used this drug in an as-needed fashion [11]. However, patients with the same genotype showed a decrease in response after regular use of albuterol. The Gly-Gly genotype at beta-2-AR-16 was unaffected by regular use. This example shows that the clinical effects of genetic variants should be interpreted in the context of patterns of use of the drug regimen over time, in particular in cases where receptor kinetics (e.g., up- and downregulation of the receptor) play a critical role. However, later clinical and epidemiologic studies, directed at optimizing asthma treatment through beta-2-AR gene information, were not able to reconfirm the clinical relevance of the earlier findings, an example of type I error (discussed later in this chapter) frequently observed in common diseases [42].

Pharmacodynamic gene–drug interactions may also result in mixed responses in terms of intended and nonintended effects. For example, the treatment of patients with schizophrenia is still unsatisfactory because of the highly variable and frequently poor response profiles of antipsychotic drugs [43]. It is thought that dopamine receptors play an important role in both achieving the wanted therapeutic benefits and the occurrence of side effects (e.g., drug-induced tardive dyskinesia and parkinsonism) with these drugs. It appears as though there is a complex interplay between available antipsychotics and an array of dopamine D2, D3, and D4 receptor actions. This example of pharmacodynamic drug–gene interactions illustrates that therapeutic responses are unlikely to be associated with a single polymorphism, in particular when the same receptor panel is responsible for both therapeutic and adverse responses.

Thus, pharmacodynamic gene–drug interactions may also affect the risk of adverse reactions. Another example is a polymorphism in the gene coding for the bradykinin B2 receptor that has been associated with an increased risk of angiotensin converting enzyme (ACE) inhibitor-induced cough [44]. Cough is one of the most frequently seen adverse drug reactions (ADRs) in ACE therapy and very often a reason for discontinuation of therapy. The TT genotype and T allele of the human bradykinin B2 receptor gene were found to be significantly higher in subjects with cough [44]. However, similar to many other studies, replication of

these findings has been limited. Further research using genome-wide association studies (GWAS) has suggested that other SNPs are related to intolerance to ACE inhibitors, but again require replication [45].

Gene–Drug Interactions and the Causal Pathway of Disease

Along with altering the pharmacokinetic and pharmacodynamic properties of medications, genetic polymorphisms may also alter the disease state that is the target of drug therapy. As an example, hypertension is widely acknowledged to be a complex phenotype that involves many regulatory systems. These regulatory systems are associated with the responsiveness to different drug therapies. Medications that work by a particular mechanism, such as the increased sodium excretion of some antihypertensive medications, may have different effects depending on the susceptibility of the patient to the effects of the drug. One key polymorphism is in the alpha-adducin gene and its relation to treatment for hypertension. Cusi *et al*. found a significant association between the alpha-adducin locus (the site of the gene) and essential hypertension and greater sensitivity to changes in sodium balance among patients with the polymorphism of the gene [46]. These findings fueled various pharmacoepidemiologic studies to evaluate whether the alpha-adducin polymorphism may also be useful to identify hypertensive patients who can optimally benefit from diuretic treatment, but with rather inconsistent results regarding the impact of the drug–gene interaction on clinical outcomes [8,47].

Genetic variability in disease states also can be critical for tailoring drug therapy to patients with a specific genotype related to both the disease and drug response. One example is the humanized monoclonal antibody trastuzumab (Herceptin®), which is used for the treatment of metastatic breast cancer patients with overexpression of the HER2 oncogene. The HER2 protein is thought to be a unique target for trastuzumab therapy in patients with this genetically associated overexpression, occurring in 10–34% of females with breast cancer [12]. The case of trastuzumab, together with another anticancer drug, imatinib, which is especially effective in patients with Philadelphia chromosome-positive leukemias, has pioneered successful genetically targeted therapy [48]. The association of somatic mutations to drug response has received substantial interest. There are many targeted therapies now available that block the growth and spread of cancer by interfering with specific molecules involved in the growth, progression, and spread of cancer.

Genetic polymorphisms that alter disease states can also play a role in drug safety. For example, factor V Leiden mutation, present in about one out of 20 Caucasians, is considered an important genetic risk factor for deep vein thrombosis and embolism [49]. A relative risk of about 30 in factor V carriers and users of oral contraceptives compared to noncarriers and non-oral contraceptive users has been reported. This gene–drug interaction has also been linked to the differential thrombotic risk associated with third-generation oral contraceptives compared with second-generation oral contraceptives [13]. Despite this strong association, Vandenbroucke *et al*. have calculated that mass screening for factor V would result in denial of oral contraceptives for about 20 000 women positive for this mutation in order to prevent one death [50]. Therefore, these authors concluded that reviewing personal and family thrombosis history and, only if suitable, factor V testing before prescribing oral contraceptives is the recommended approach to avoid this adverse gene–drug interaction [50]. This highlights another important role of molecular pharmacoepidemiology: determining the utility and cost-effectiveness (see also Chapter 34) of genetic screening to guide drug therapy [51].

The Interplay of Various Mechanisms

It is useful to conceptualize how the effects of genetic polymorphisms at different stages of drug disposition and response might influence an individual's response to a medication. As an example, an individual may have a genotype that alters the metabolism of the drug, the receptor for the drug, or both [25]. Depending on the combination of these genotypes, the individual might have a different response in terms of both efficacy and toxicity (Table 30.1). In the simplified example in Table 30.1, there is one genetic variant that alters drug metabolism and one genetic variant that alters receptor response to a medication of interest. In this example, among those who are homozygous for the alleles that encode normal drug metabolism and normal receptor response, there is relatively high efficacy and low toxicity. However, among those who have a variant that reduces drug metabolism, efficacy at a standard dose could actually be greater (assuming a linear dose–response relationship within the possible drug levels of the medication) but toxicity could be increased (if dose related). Among those who have a variant that reduces receptor response, drug efficacy will be reduced while toxicity may not be different from those who carry genotypes that are not associated with impaired receptor response (assuming that toxicity is not related to the receptor responsible for efficacy). Among those who have variants for both genes, efficacy could be reduced because of the receptor variant (perhaps not as substantially as those with an isolated variant of the receptor gene because of the higher effective dose resulting from the metabolism gene variant), while toxicity could be increased because of the metabolism variant.

Some Examples of the Progression and Clinical Application of Molecular Pharmacoepidemiology

Medications with a narrow therapeutic ratio are good targets for the use of molecular pharmacoepidemiology to improve the use and application of medications. One example is warfarin. This example illustrates both the logical progression of pharmacogenetics through molecular pharmacoepidemiology and the complexity of moving pharmacogenetic data into practice. The enzyme primarily responsible for the metabolism of warfarin to its inactive form is the cytochrome P450 2C9 variant (CYP2C9) [52–54]. Pharmacogenetic studies identified polymorphisms in CYP2C9 that led to altered metabolism of warfarin [55,56].

Table 30.1 Hypothetical response to medications by genetic variants in metabolism and receptor genes.

		Drug response	
Gene affecting metabolism*	**Gene affecting receptor response***	**Efficacy**	**Toxicity**
Wild type	Wild type	70%	2%
Variant	Wild type	85%	20%
Wild type	Variant	20%	2%
Variant	Variant	35%	20%

Modified from Evans and McLeod [25].
*Wild type associated with normal metabolism or receptor response and variants associated with reduced metabolism or receptor response.

One of the first molecular pharmacoepidemiologic studies examining the clinical relevance of the CYP2C9 variants was a case–control study that reported that the odds ratio (OR) for a low warfarin dose requirement was 6.2 (95% CI 2.5, 15.6) among those having one or more CYP2C9 variant alleles compared with a control population with normal warfarin dose requirements [57]. The OR was elevated both in those with only one variant allele (i.e., heterozygotes: OR 2.7; 95% CI 1.2, 5.9) and in those with two variant alleles to an even greater extent (i.e., homozygotes: OR 7.8; 95% CI 1.9, 32.1). Patients on low doses of warfarin also were more likely to have difficulty with anticoagulation control during the first week of therapy and more likely to have bleeding complications, based on unadjusted analyses. A subsequent retrospective cohort study confirmed the lower dose requirement of patients with the genetic variant of CYP2C9, but did not examine clinical outcomes [58].

In order to address the clinically relevant question of bleeding, another retrospective cohort study was performed that demonstrated an increased risk of bleeding among patients followed in an anticoagulation clinic who had at least one variant of the CYP2C9 genotype [59]. The relatively small size of the study, retrospective nature, and selected population left unanswered the question of whether there is an independent effect of CYP2C9 variants on the risk of clinical outcomes throughout the course of anticoagulation therapy, whether specific variants or combinations of variants (e.g., heterozygotes with only one variant allele versus homozygotes with two variant alleles) have different effects, and whether knowing that a patient carries a variant can alter therapy in a way that can reduce risk.

A metaanalysis of studies examining the role of CYP2C9 in warfarin-treated patients demonstrated a significant association between CYP2C9 variants and bleeding risk [60]. Additional research clearly demonstrated that the vitamin K epoxide reductase complex 1 (VKORC-1) gene carries several variants that alter response to warfarin. Of note, most of the strongest associations with warfarin dose are among variants that are all in strong linkage disequilibrium with each other, particular in non-African American populations; thus, there is no benefit to dose prediction in these patients in genotyping more than just one SNP. Despite the presence of two genes with relatively strong associations with warfarin dosing, there is still about 50% of variability in warfarin dosing that is not explained by genetics or clinical factors, suggesting that other genetic factors may also influence the response to the medication [61]. However, despite much research, very few other SNPs have been identified that have a substantial effect on warfarin dosing, suggesting that perhaps many variants, including other variants in CYP2C9 and VKORC1 that may be more important in African-Americans, each with only a relatively small effect on dose, may be needed to add to our ability to predict warfarin response.

Eventually, clinical trials were performed to answer the important question of clinical utility: does altering warfarin dosing based on genotype affect outcomes? Following smaller, nondefinitive trials, three large-scale clinical trials of warfarin have been conducted. The first two had very different designs. The Clarification of Optimal Anticoagulation through Genetics (COAG) trial had greater control of dose titration in both study arms and used a comparison group that differed from the intervention group only in the absence of the use of genetics [62]. The EU-PACT UK trial used a formal dosing algorithm but left dose titration up to the discretion of practitioners who were not blinded to dose [63]. Perhaps most importantly, the comparison group was a usual care group and therefore did not incorporate a formal clinical-only (i.e., without genetics) initial dosing algorithm. In addition, while COAG enrolled about 27% African-Americans, EU-PACT UK enrolled only 1%. The two trials thus answered different

questions. COAG addressed whether adding genetics to clinical factors in a formal dosing algorithm can improve anticoagulation control relative to using a clinical-only algorithm under uniform dose titration methods and blinding of study subjects and practitioners. EU-PACT UK addressed whether a formal algorithm that used both genetic and clinical information improves anticoagulation relative to fixed-dose initiation where the dose titration method was left up to unblended clinicians. The COAG trial demonstrated no benefit of pharmacogenetic dosing on anticoagulation control overall, and worsening of anticoagulation control with pharmacogenetic dosing in African-Americans. The EU-PACT UK trial demonstrated improvement in anticoagulation control.

A third trial, the GIFT trial, examined pharmacogenetic dosing in orthopedic patients, compared with a clinical algorithm dosing arm [64]. This trial demonstrated benefit from pharmacogenetic dosing, but could not address the question of the effects of pharmacogenetics in African-Americans due to the enrollment of very few African-American patients.

Together, these trials demonstrate the need for, and benefit of, pharmacogenetic clinical trials testing different strategies in different patient populations.

Another pertinent example is in oncology. Cancer is an extremely heterogeneous disease with differences not only between cancer cells from different patients but also between cancer cells within a single patient. Every cancer patient exhibits a different genetic profile and the profile can change over time; thus, more patients will benefit if therapeutic options can be tailored to each individual, thus avoiding the "one size fits all" cancer treatment. Markers predicting response to anticancer drugs are mostly related to the fact that drug efficacy can be greatly influenced by alterations in drug targets and in related proteins present in tumor cells. Therefore, cancer-targeted therapies, directed to a specific cancer alteration, may only be indicated for the subgroup of patients with tumors carrying that molecular target. Examples include trastuzumab and imatinib mentioned earlier in the chapter.

To date, there is information about predictive biomarkers for many of the approved oncology drugs; the number of predictive biomarkers in cancer exceeds that in any other medical field. The US FDA currently includes over 300 pharmacogenetic labels, and a third of these are related to oncology. Most oncology-related markers indicate that drug use should be tailored to a subgroup of patients who have the targeted molecular alteration that predict drug efficacy (e.g., crizotinib for ALK rearrangements; bosutinib, omacetaxine, and ponatinib for BCR–ABL fusion protein; and dabrafenib, trametinib, and vemurafenib for mutated BRAF). The EMA also provides information regarding pharmacogenomic biomarkers in drug labels. In addition, several working groups, such as the PharmGKB (www.pharmgkb.org/view/drug-labels.do) and the National Comprehensive Cancer Centre (NCCN) Task Force, have published guidelines for the implementation of biomarkers in clinical practice to guide selection of drug treatments. Recommendations by several groups regarding the validity and utility of genetic tests for clinical practice are also available, for example by the Centers for Disease Control and Prevention (CDC) Office of Public Health Genomics (https://phgkb.cdc.gov/PHGKB/topicStartPage.action), which ranks genomic tests according to evidence level, and the Clinical Utility Gene Cards (CUGCs), which are disease-specific guidelines regarding the clinical utility of genetic testing authored by international experts (www.eurogentest.org/index.php?id=668).

It is now recognized that tumors have unique molecular compositions and can be subdivided accordingly. Each unique tumor type might be most sensitive to a specific drug combination. By identifying key driver mutations, specific drugs can be matched to the molecular targets

in the tumor and provide a "personalized" cancer treatment. While therapies for single targets are still common (e.g., trastuzumab targeting HER2 in treatment of breast cancer), use of therapies directed to multiple targets (e.g., sorafenib and sunitinib targeting the vascular endothelial growth factor receptor family [VEGF], platelet-derived growth factor receptor family [PDGFR], fms-like tyrosine kinase 3 [FLT-3] and stem cell factor receptor c-Kit in treatment of nonsmall cell lung cancer) are likely to increase to avoid relapses that occur frequently with targeted therapies. Precision oncology is enabled by technological advances. Next-generation sequencing technologies (described later in the chapter) can be applied to formalin-fixed paraffin-embedded tumor tissues (the samples routinely available in the clinic) to identify a large number of clinically relevant alterations in a timely and cost-effective manner and prospectively select cancer treatment.

It is important to note that novel treatment strategies can be based on a specific genetic characteristic regardless of the type or subtype of cancer. For instance, in 2017 the US FDA approved pembrolizumab for treatment of unresectable or metastatic solid tumors that have a biomarker, microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR), in adult and pediatric patients. This is the first drug approval based on a tumor's biomarker without regard to the tumor's original location.

Methodologic Problems to be Addressed by Pharmacoepidemiologic Research

As previously discussed, the basic science of molecular pharmacoepidemiology is the same basic science underlying pharmacoepidemiology. Therefore, the same methodologic problems of pharmacoepidemiology must be addressed in molecular pharmacoepidemiology. These problems include those of chance and statistical power, confounding, bias, and generalizability (see Chapters 3, 4, and 43).

However, the complex relationship between medication response and molecular and genetic factors generates some unique challenges in molecular pharmacoepidemiology. Many of these challenges derive from the large number of potential genetic variants that can modify the response to a single drug, the possibility that there is a small individual effect of any one of these genes, the low prevalence of many genetic variants, and the possibility that a presumptive gene–drug response relationship may be confounded by the racial and ethnic mixture of the population studied [65,66]. Thus, the methodologic challenges of molecular pharmacoepidemiology are closely related to issues of statistical interactions, type I and type II errors, and confounding.

First and foremost, however, molecular pharmacoepidemiologic studies rely on proper identification of putative genes. In addition, in all research of this type, use of appropriate laboratory methods, including the use of high-throughput genotyping technologies, is necessary. Similarly, appropriate quality control procedures must be considered to obtain meaningful data for research and clinical applications. Recent next-generation sequencing techniques have only highlighted further the need for, and complexity of, obtaining valid genotyping results. This section will focus on the methodologic challenges of studying interactions, minimizing type I and type II errors, and accounting for confounding, particularly by population admixture (defined below).

Interactions

Along with examining the direct effect of genes and other biomarkers on outcomes, molecular pharmacoepidemiologic studies must often be designed to examine effect modification between medication use and the genes or

biomarkers of interest. That is, the primary measure of interest is often the role of biomarker information on the effect of a medication. For purposes of simplicity, this discussion will use genetic variability as the measure of interest.

Effect modification is present if there is a difference in the effect of the medication depending on the presence or absence of the genetic variant. This difference can be either on the multiplicative or additive scale. On the multiplicative scale, interaction is present if the effect of the combination of the genotype and medication exposure relative to neither is greater than the product of the measure of effect of each (genotype alone or medication alone) relative to neither. On the additive scale, interaction is present if the effect of the combination of the genotype and medication exposure is greater than the sum of the measures of effect of each alone, again all relative to neither [67].

For studies examining a dichotomous medication exposure (e.g., medication use versus nonuse), a dichotomous genetic exposure (e.g., presence versus absence of a genetic variant), and a dichotomous outcome (e.g., myocardial infarction occurrence versus none), there are two ways to consider presenting and analyzing interactions [68]. The first is as a stratified analysis, comparing the effect of medication exposure versus nonexposure on the outcome in two strata: those with the genetic variant and those without (for example, see Table 30.2). The second is to present a 2 × 4 table (also shown in Table 30.2). In the first example (stratified analysis), one compares the effect of the medication among those with the genetic variant to the effect of the medication among those without the genetic variant. In the second example (the 2 × 4 table), the effect of each combination of exposure (i.e., with both genetic variant and medication; with genetic variant but without medication; with medication but without genetic variant) is determined relative to the lack of exposure to either. The advantage of the 2 × 4 table is that it presents separately the effect of the drug, the gene, and both relative to those without the genetic variant and without medication exposure. In addition, presentation of the data as a 2 × 4 table allows one to directly

Table 30.2 Two ways to present effect modification in molecular pharmacoepidemiologic studies using case–control study as a model.

Genotype	Medication	Cases	Controls	Odds ratio	Information provided
Stratified analysis					
+	+	a	b	ad/bc	Effect of medication vs no medication among those with the genotype
	-	c	d		
-	+	e	f	eh/fg	Effect of medication vs no medication among those without the genotype
	-	g	h		
2 × 4 table					
+	+	a	b	ah/bg = A	Joint genotype and medication vs neither
+	-	c	d	ch/dg = B	Genotype alone vs neither
-	+	e	f	eh/fg = C	Medication alone vs neither
-	-	g	h	Reference	Reference group

Modified from Khoury MJ, Little J, Burke W, eds. *Human Genome Epidemiology*. New York: Oxford University Press, 2004.

compute both multiplicative and additive interactions [68].

In the example given in Table 30.2, multiplicative interaction would be assessed by comparing the odds ratio for the combination of genotype and medication exposure to the product of the odds ratios for medication alone and genotype alone. Multiplicative interaction would be considered present if the odds ratios for the combination of medication and genotype (A in Table 30.2) was greater than the product of the odds ratios for either alone (B × C). Additive interaction would be considered present if the odds ratio for the combination of genotype and medication use (A) was greater than the sum of the odds ratios for medication use alone and genotype alone (B + C). The 2 × 4 table also allows the direct assessment of the number of subjects in each group along with the respective confidence interval for the measured effect in each of the groups, making it possible to directly observe the precision of the estimates in each of the groups and therefore better understand the power of the study. Furthermore, attributable fractions can be computed separately for each of the exposures alone and for the combination of exposures.

In general, we believe that presenting the data in both manners is optimal because it allows the reader to understand the effect of each of the exposures (2 × 4 table) as well as the effect of the medication in the presence or absence of the genotypic variant (stratified table).

Type I Error

The chance of type I error (concluding there is an association when in fact one does not exist) increases with the number of statistical tests performed on any one data set (see also Chapter 4) [69]. It is easy to appreciate the potential for type I error in a molecular pharmacoepidemiologic study that examines, simultaneously, the effects of multiple genetic factors, the effects of multiple nongenetic factors, and the interaction between and among these factors [69–71]. One of the reasons cited for nonreplication of study findings in molecular pharmacoepidemiology is type I error [42]. Limiting the number of associations examined to those of specific candidate genetic variants that are suspected of being associated with the outcome is one method to limit type I error in pharmacoepidemiology [72]. However, with increasing emphasis in molecular pharmacoepidemiologic studies on identifying all variants within a gene (and all variants within the genome) and examining multiple interactions, this method of limiting type I error is often not tenable [73]. Some other currently available solutions are discussed in the next section.

Type II Error

Because it has been hypothesized that much of the genetic variability leading to phenotypic expression of complex diseases results from the relatively small effects of many relatively low-prevalence genetic variants [74], the ability to detect a gene–response relationship is likely to require relatively large sample sizes to avoid type II error (concluding there is no association when in fact one does exist) (see also Chapter 4) [75]. The sample size requirements for studies that examine the direct effect of genes on medication response will be the same as the requirements for examining direct effects of individual risk factors on outcomes. With relatively low prevalences of polymorphisms and often low incidence of outcomes (particularly in studies of adverse drug reactions), large sample sizes are typically required to detect even modest associations. For such studies, the case–control design (see Chapter 3) has become a particularly favored approach for molecular pharmacoepidemiologic studies because of its ability to select participants based on the outcome of interest (and its ability to study the effects of multiple potential genotypes in the same study).

Studies designed to examine the interaction between a genetic polymorphism and a medication will require even larger sample sizes [76]. This is because such studies need to be powered to compare those with both the genetic polymorphism and the medication exposure with those who have neither. As an example, the previously mentioned case–control study of the alpha-adducin gene and diuretic therapy in patients with treated hypertension examined the effects of the genetic polymorphism, the diuretic therapy, and both in combination [8]. There were a total of 1038 participants in the study. When comparing the effect of diuretic use with no use and comparing the effect of the genetic variant with the nonvariant allele, all 1038 participants were available for comparison (Table 30.3). However, when examining the effect of diuretic therapy versus nonuse among those with the genetic variant, only 385 participants contributed to the analyses. Of note, this study presented the data for interaction in the two ways presented in Table 30.2.

In order to minimize false-negative findings, further efforts must be made to ensure adequate sample sizes for molecular pharmacoepidemiologic studies. Because of the complex nature of medication response, and the likelihood that at least several genes are responsible for the variability in drug response, studies designed to test for multiple gene–gene and gene–environment interactions (including other medications, environmental factors, adherence to medications, and clinical factors) will, similarly, require large sample sizes.

Confounding by Population Admixture

When there is evidence that baseline disease risks and genotype frequencies differ among ethnicities, the conditions for population stratification (i.e., population admixture or confounding by ethnicity) may be met [77]. Population admixture is simply a manifestation of confounding by ethnicity, which can occur if both baseline disease risks and genotype frequency vary across ethnicity.

For example, the African-American population represents admixture of at least three major continental ancestries (African, European, and Native American). Wacholder *et al.* demonstrated that the larger the number of ethnicities involved in an admixed population, the less

Table 30.3 Gene–exposure interaction analysis in a case–control study.

Diuretic use	Adducin variant	Cases	Controls	Odds ratio (OR) for stroke or myocardial infarction
0	0	A_{00} 103	B_{00} 248	1.0
0	1	A_{01} 85	B_{01} 131	1.56
1	0	A_{10} 94	B_{10} 208	1.09
1	1	A_{11} 41	B_{11} 128	0.77

Case control OR in variant carriers: $OR_{variant} = A_{11}B_{01}/A_{01}B_{11} = 41 \times 131/85x128 = 0.49$
Case control OR in wild-type carriers: $OR_{wild\text{-}type} = A_{10}B_{00}/A_{00}B_{10} = 94 \times 248/103 \times 208 = 1.09$
Synergy index = $OR_{variant}/OR_{wild\text{-}type} = 0.45$
Case-only OR = $A_{11}A_{00}/A_{10}A_{01} = 41 \times 103/94 \times 85 = 0.53$
Adapted from Psaty *et al.* [8].

likely that population stratification can be the explanation for biased associations [77]. Millikan [78] and Wang *et al*. [79] also reported that a minimal bias in point estimates is likely in African-American populations, suggesting that point estimates of association will not usually be influenced by population stratification in studies that involve African-American populations under most usual circumstances. Ardlie *et al*. used empirical data to show that carefully matched, moderate-sized case–control samples in African-American populations are unlikely to contain levels of population admixture that would result in significantly inflated numbers of false-positive associations [80]. They did observe the potential for population structure to exist in African-American populations, but this structure was eliminated by removing recent African or Caribbean immigrants, and limiting study samples to resident African-Americans. Furthermore, Cardon and Palmer argued that poor study design may be more important than population stratification in conferring bias to association studies [81].

Based on the literature evaluating the effects of confounding by ethnicity overall, and specifically in African-Americans, there is little empirical evidence that population stratification is a likely explanation for bias in point estimates or incorrect inferences [77]. Nonetheless, population admixture must be considered in designing and analyzing molecular pharmacoepidemiologic studies to ensure that adequate adjustment can be made for this potential confounder. New approaches to addressing population admixture are presented in the following section.

Currently Available Solutions

Identifying Additional Genetic Contributors to Drug Response

A great concern of the identification of low-penetrance alleles is that they have not yet been able to explain the majority of the estimated genetic contribution to disease etiology. Based on studies of families or phenotypic variability, most loci have been found to explain less than half (and at times as little as 1%) of the predicted heritability of many common traits [82]. This "missing heritability" of complex disease suggests that other classes of genetic variation may explain much of the genetic contribution to common disease.

There currently are two primary approaches for gene discovery: candidate gene association studies and genome-wide studies. In the former, genes are selected for study on the basis of their plausible biological relevance to drug response. While this allows for identification of variants with *a priori* biological plausibility, it is limited by our partial knowledge of which genetic variants may actually be responsible for variable drug effects. In the latter, DNA sequences are examined for associations with outcomes, initially irrespective of biological plausibility. The benefit of this approach is that it does not rely on our limited knowledge of genetics; the disadvantage is that the biological plausibility of the findings may then need to be confirmed.

One example is GWAS which rely on LD, defined above as the correlation between alleles at two loci. This approach uses DNA sequence variation (e.g., SNPs) found throughout the genome, and does not rely on *a priori* functional knowledge of gene function. Therefore, GWAS can be used to identify new candidate genes or regions, but relies on the potential for truly causative gene effects to be detected using genetic variants that may not have a functional effect. A number of factors influence the success of these studies. Appropriate epidemiologic study designs and adequate statistical power remain essential. Thorough characterization of LD is essential for replication of GWAS: the haplotype mapping (HapMap) consortium and other groups have shown that the extent of LD varies by ethnicity, which may affect the ability to replicate findings in subsequent studies [74]. Particularly informative SNPs that best

characterize a genomic region can be used to limit the amount of laboratory and analytical work in haplotype-based studies [83]. It has been hypothesized that studies that consider LD involving multiple SNPs in a genomic region (i.e., a haplotype) can increase power to detect associations by 15–50% compared with analyses involving only individual SNPs [84]. Finally, even if genome-wide scans may identify markers associated with the trait of interest, a challenge will be to identify the causative SNPs.

Newer, sequencing technologies have made it possible to study rarer genetic variants. While Sanger sequencing is still considered the gold standard in clinical testing, its limitations include low throughput and high cost. Broadly, next-generation sequencing (NGS) describes technologies that utilize clonally amplified or single-molecular templates that are then sequenced in a massively parallel fashion. The advance of NGS technologies has been enabled by innovation in sequencing chemistries, better imaging, microfabrication and information technology. In addition, bioinformatics tools for data analysis and management and sample preparation methods have rapidly evolved along with the sequencing technologies, translating to reductions in the amount of input materials required. In 2013, the US FDA approved marketing for the first time for a next-generation sequencer, Illumina's MiSeqDx, which allows the development and use of innumerable new genome-based tests.

Clearly, candidate gene and genome-wide approaches are not mutually exclusive. Both have the potential to identify important variants that may be clinically useful.

Interactions

Along with traditional case–control and cohort studies, the case-only study can be used for molecular pharmacoepidemiologic studies designed to examine interactions between genes and medications [85,86]. In this design, cases, representing those with the outcome or phenotype of interest, are selected for study, and the association between genetic variants and medication use is determined among these cases. Under the assumption that there is no association between the gene and medication exposure among those without the disease (i.e., controls), the odds ratio for the association between genetic variants and medication use in the cases is equivalent to the synergy index on a multiplicative scale for a case–control study [68]. (The synergy index is the odds ratio for medication use versus the outcome of interest in those with the variant alleles divided by the odds ratio for medication use versus the outcome in those without the variant alleles – see Table 30.3 footnote.) In other words, assuming that the use of the medication is unrelated to the genotype, the case-only study provides a valid measure of the interaction of the genotype and the medication on the risk of the outcome.

One strength of the case-only study design is that it eliminates the need to identify controls, which is often a major methodologic and logistical challenge in case–control studies. In addition, the case-only study can result in greater precision in estimating interactions compared with case–control analyses [85,86]. It also is possible to use the case-only approach to estimate interactions between genes and medications in large-scale registries of people with diseases or disease outcomes (e.g., cancer registries with genotypes and medication information available) [68].

One example of a case-only study is a study examining antihypertensive medication pharmacogenetics [87]. By using a large-scale clinical trial database derived from the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), the authors examined cases of coronary heart disease and cases of heart failure outcomes and tested for gene-by-treatment effects among the cases. They

identified a potential interaction of a polymorphism within the ryanodine receptor 3 gene and heart failure outcomes.

There are several limitations of the case-only design [85]. As stated above, the design relies on the assumption of independence between exposure (medication use) and genotype. Although this assumption may be valid (in the absence of knowing genotype clinically, it may be reasonable to assume that the use of the medication is not related to patients' genotypes), it is certainly possible that, within observational studies, the genotype, by altering response to medications targeted at a specific disease or by altering the disease, could affect the medications being prescribed to patients. For example, the use of a particular antihypertensive medication may be related to prior success with other medications. Patients carrying genotypic variants that diminish the response to one class of antihypertensive medication may be more likely to be on other classes of antihypertensive medications. Thus, there would be an association between the genotype and the medication exposure. One way to minimize this possibility is to include only first-time prescriptions for hypertensive medications. Another method is to perform the case-only study within a randomized trial, where drug use is randomly assigned, as in the ALLHAT example provided above.

Another limitation of the case-only design is that it does not allow assessment of the independent effects of medication use or genotype on outcome. Further, the assessment of interaction can only be interpreted on a multiplicative scale.

Type I Error and Replication

Given concerns of type I error (along with other methodologic concerns such as uncontrolled confounding, publication bias, and linkage disequilibrium), a key issue in molecular epidemiology is the ability to replicate association study findings. Replication of association studies is required not only to identify biologically plausible causative associations, but also to conclude that a candidate gene has a meaningful etiologic effect. Lohmueller *et al.* observed that many associations are not replicated [42]. This lack of replication can be explained by false-positive reports (e.g., spurious associations), false-negative reports (e.g., studies that are insufficiently powerful to identify the association), or actual population differences (e.g., the true associations are different because of differences in genetic background, exposures, etc.). Given the perceived lack of consistency in association studies, what level of confidence can we have in associations reported to date?

Lohmueller *et al.* addressed these issues by undertaking a metaanalysis of 25 inconsistent associations and 301 "replication" studies (i.e., by ignoring the initial positive report) [42]. Most initial associations were not replicated, but an excess (20%) of replicated associations was seen, while only 5% were expected under the null hypothesis. This replication is not solely due to publication bias, since one would have to hypothesize that 40–80 negative studies were not reported rather than the average of 12 reported studies per association. Lohmueller *et al.* also concluded that it was unlikely that these replications represented false positives due to ethnic stratification. Different linkage disequilibrium patterns or other population patterns or population-specific modifiers (genes and/or environments) could also explain lack of replication, but this was unlikely to be a significant source of study inconsistency. The first positive reports also tended to be unreliable estimates for subsequently reported ORs [88], perhaps due to the "winner's curse" phenomenon which predicts that the initial positive report overestimates the "true" value [89]. Indeed, 23 of 25 associations studied showed evidence for a "winner's curse." An additional consequence of this phenomenon is

that replication studies may therefore require larger sample sizes since the actual effects being replicated may be smaller than suggested by the initial report.

Despite these limitations, these data indicate that associations are replicable more often than expected by chance, and may therefore represent truly causative effects on disease. Nonetheless, as illustrated by several examples in this chapter, replication is not guaranteed, and one cannot assume that an association from a single association is real.

In order to achieve believable, replicable association results, investigators must consider factors that influence the design, analysis, and interpretation of these studies. These include, as discussed above, adequate sample size, proper study design, and characterization of the study population, particularly when replication studies themselves are not comparable in terms of ethnicity or other confounding factors.

Data analytical methods can complement replication studies to address multiple testing and type I error, which are common problems in pharmacogenetics studies evaluating multiple SNPs, multiple exposures, and multiple interactions. Bonferroni correction is the most basic approach for adjusting multiple testing but is considered too conservative for tightly linked SNPs and may wipe out many small effects that one may actually expect (increased risk of type II errors). The false discovery rate (FDR) approach is less conservative for controlling for multiple analyses of the data. The FDR method estimates the expected proportion of false positives among associations that are declared significant, which is expressed as a q-value [90]. Under a Bayesian approach, there is no penalty for multiple testing because the prior probability of an association should not be affected by the tests that the investigator chooses to conduct. However, without strict standards, investigators might be tempted to cut corners or exaggerate the prior plausibility of a model that is supported *a posteriori* [91].

Type II Error

Reducing type II error essentially involves a logistical need to ensure adequate sample size (see also Chapter 4). One approach to increasing the sample size of molecular pharmacoepidemiologic studies is to perform large, multicenter collaborative studies. Another is to combine multiple, separately performed cohorts, sometimes referred to as metaepidemiologic studies. One example was the International Warfarin Pharmacogenetics Consortium (IWPC). This consortium of over 21 centers across nine countries combined data from multiple cohort studies in order to develop multiethnic dosing algorithms, attempt to identify uncommon SNPs associated with warfarin response, and perform GWAS [92]. By combining cohorts, the IWPC became the largest sample size of any warfarin pharmacogenetics studies.

Another potential solution to minimizing type II error is through metaanalysis, whereby smaller studies, which are individually not powered to detect specific associations (such as interactions), are combined in order to improve the ability to detect such associations (see Chapter 36). One particularly intriguing approach is the concept of prospective meta-analysis in which studies are planned or identified in advance of performing a metaanalysis so that important elements of study design complement each other across studies and important potential sources of bias that hamper the interpretation of retrospective metaanalyses can be avoided (see Chapter 36).

Adequate reporting of genetic association studies is important to allow assessment of their strengths and weaknesses. The STREGA statement (Strengthening the Reporting of Genetic Association studies) is an extension of the STROBE statement (Strengthening the Reporting

of Observational Studies in Epidemiology) that provides a checklist to help researchers and journals [93].

Population Admixture

As presented above, although population stratification is unlikely to be a significant source of bias in epidemiologic association studies, this assumes adequate adjustment for race. A number of analytical approaches exist to either circumvent problems imposed by population genetic structure or use this structure in gene identification [94,95]. The "structured association" approach identifies a set of individuals who are drawing their alleles from different background populations or ethnicities. This approach uses information about genotypes at loci that lie in regions other than the location of the gene of interest (i.e., "unlinked markers") to infer their ancestry (often referred to as ancestry informative markers) and learn about population structure. It further uses the data derived from these unlinked markers to adjust the association test statistic. By adjusting for these ancestry informative markers, one can adjust for differences in ancestry. The ALLHAT study, described above, performed such an adjustment using ancestry informative markers.

The Future

Without any doubt, scientific and clinical developments in biology and (bio)chemistry, particularly in the field of genomics and other biomarkers, have affected and will continue to affect the field of pharmacoepidemiology in a significant way. As discussed earlier in this chapter, translating biomarkers from the lab and experimental studies to clinical practice, and thereby to the study field of molecular pharmacoepidemiology, has been a difficult path.

We have addressed in this chapter several examples where the initial promising findings on drug–gene interactions to predict clinical drug responses could not be replicated in subsequent studies. For sure, the ability of genes and other biomarkers to improve patient care and outcomes will need to be tested in properly controlled studies, including randomized controlled trials. The positive and negative predictive value of carrying a genetic variant will be important determinants of the ability of the variant to improve outcomes. Those genetic variants with good test characteristics may still need to be evaluated in properly controlled trials. Such studies could examine several ways to incorporate genetic testing into clinical practice, including the use of genetic variants in dosing algorithms [96,97], in selection of a specific therapeutic class of drug to treat a disease [8], and in avoidance of using specific medications in those at high risk for adverse drug reactions [44]. These scientific advances are also finding their way into drug discovery and development in order to rationalize drug innovation and to identify good and poor responders, both in terms of efficacy and safety, to drug therapy in an earlier phase [98].

The cost-effectiveness of such approaches is also of great interest because the addition of genetic testing adds cost to clinical care (see also Chapter 34). Veenstra and colleagues have developed a set of criteria for evaluating the potential clinical and economic benefits of pharmacogenetic testing [99]. These criteria include the severity of the outcome avoided, the availability of other means to monitor drug response without the need for additional pharmacogenetics testing, the strength of the association between the genetic variants and clinically relevant outcomes, the availability of a rapid and relatively inexpensive assay, and the frequency of the variant alleles. In essence, these criteria could be applied to any new diagnostic test. Clearly, additional research will be needed to determine the cost-effectiveness of new biomarker and genetic tests as they are developed.

Next-generation sequencing will also require the development of novel approaches to data analyses. There are three levels of analysis that are conducted by NGS technologies: (i) targeted gene panels focus on a limited set of genes allowing for greater depth of coverage. The advantages include higher analytical sensitivity and specificity, and improved ability to interpret the results in a clinical context because only genes with an established role in the disease are sequenced; (ii) exome sequencing tests all coding regions of the human genome; and (iii) whole-genome sequencing analyzes the entire 3 billion bases of the genome. The targeted approach to genome sequencing is the more widespread clinical implementation of NGS technologies. This is because only some of the enormous amount of genetic information generated by exome or whole-genome sequencing can be interpreted and is actionable. Along with the bioinformatics challenges of managing and validating such large datasets, a significant amount of information will be novel and/or of unknown clinical importance.

A major area that requires further development is establishing the clinical utility of the identified markers/strategies for patients and healthcare systems. The level of evidence required to establish that a marker is clinically useful and should be introduced for routine use has been discussed extensively but consensus has not been reached.

Nevertheless, genetic and molecular studies are increasingly being incorporated in large clinical trials, which can lead to the identification of subgroups of patients with clear benefit from drugs, accelerating the discovery of effective therapies for selected populations. Another challenge to the implementation of genetic testing is the fact that pharmacogenetics knowledge is constantly being updated. Clinicians need to interpret the results of these tests in accordance with current understanding of the association between pharmacogenetic variation and drug effects.

What this all means for the future of pharmacoepidemiology is a challenging question. Genotype data will increasingly become available and will enrich pharmacoepidemiologic analysis. New methods (e.g., sequencing) will provide new opportunities, but also new challenges to analyzing pharmacoepidemiologic data. Further, although it is useful to characterize the three different pathways of how drug–gene interactions may occur, as was done in this chapter, this stratification is most likely an oversimplification of the large plethora of possible mechanisms of how drugs, genes, and patient outcomes are interrelated. All these may have consequences for how molecular pharmacoepidemiologic studies are designed, conducted, and analyzed. In addition, the more genotype testing is applied in clinical practice, the more drug exposure will be influenced by such tests, making genotype and drug exposure nonindependent factors.

Finally, just as for all research, the ethical, legal, and social implications of genetic testing must be considered and addressed [5,100–102] (see also Chapter 31). Pharmacogenetic testing raises issues of privacy concerns, access to healthcare services, and informed consent. For example, concern has been raised that the use of genetic testing could lead to targeting of therapies to only specific groups (ethnic or racial) of patients, ignoring others, and to loss of insurance coverage for certain groups of individuals [102]. There also is a concern that medicines will be developed only for the most common, commercially attractive, genotypes, leading to "orphan genotypes" [103,104]. Equally importantly, as more and more genetic data are collected on individuals as part of routine clinical care, the requirements and methods for returning unanticipated genetic results must be carefully determined and implemented.

All of these issues are challenges to overcome as we continue to reap the benefits of the tremendous strides made in determining the molecular basis of disease and drug response.

References

1 Reidenberg MM. Evolving ways that drug therapy is individualized. *Clin Pharmacol Therapeut* 2003; **74**: 197–202.

2 Maitland-van der Zee AH, de Boer A, Leufkens HG. The interface between pharmacoepidemiology and pharmacogenetics. *Eur J Pharmacol*. 2000; **410**: 121–30.

3 Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics* 1998; **8**: 283–9.

4 Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *JAMA* 2001; **286**: 2270–9.

5 Johnson JA. Pharmacogenetics: potential for individualized drug therapy through genetics. *Trends Genet* 2003; **19**: 6606.

6 Roden DM. Cardiovascular pharmacogenomics. *Circulation* 2003; **108**: 3071–4.

7 Roden DM, George AL Jr. The genetic basis of variability in drug responses. *Nature Rev Drug Discov* 2002; **1**: 37–44.

8 Psaty BM, Smith NL, Heckbert SR, *et al*. Diuretic therapy, the alpha-adducin gene variant, and the risk of myocardial infarction or stroke in persons with treated hypertension. *JAMA* 2002; **287**: 1680–9.

9 Rogers JF, Nafziger AN, Bertino JS Jr. Pharmacogenetics affects dosing, efficacy, and toxicity of cytochrome P450-metabolized drugs. *Am J Med* 2002; **113**: 746–50.

10 Krynetski E, Evans WE. Drug methylation in cancer therapy: lessons from the TPMT polymorphism. *Oncogene* 2003; **22**: 7403–13.

11 Israel E, Drazen JM, Liggett SB, *et al*. The effect of polymorphisms of the beta(2)-adrenergic receptor on the response to regular use of albuterol in asthma. *Am J Respir Crit Care Med* 2000; **162**: 75–80.

12 Slamon DJ, Leyland-Jones B, Shak S, *et al*. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001; **344**: 783–92.

13 Kemmeren JM, Algra A, Meijers JC, *et al*. Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor V Leiden mutation: a randomized trial. *Blood* 2004; **103**: 927–33.

14 Alving AS, Carson PE, Flanagan CL, Ickes CE. Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science* 1956; **124**: 484–5.

15 Kalow W. Familial incidence of low pseudocholinesterase level. *Lancet* 1956; **268**: 576–7.

16 Evans DA, Manley KA, McKusick VA. Genetic control of isoniazid metabolism in man. *BMJ* 1960; **5197**: 485–591.

17 Sindrup SH, Brosen K. The pharmacogenetics of codeine hypoalgesia. *Pharmacogenetics* 1995; **5**: 335–46.

18 Cargill M, Altshuler D, Ireland J, *et al*. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; **22**: 231–8.

19 Barnes MR, Breen G. A short primer on the functional analysis of copy number variation for biomedical scientists. *Methods Molec Biol* 2010; **628**: 119–35.

20 Enright EF, Gahan CG, Joyce SA, Griffin BT. The impact of the gut microbiota on drug metabolism and clinical outcome. *Yale J Biol Med* 2016; **89**: 375–82.

21 Paul B, Barnes S, Demark-Wahnefried W, *et al*. Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases. *Clin Epigenet* 2015; **7**: 112.

22 Cascorbi I, Schwab M. Epigenetics in drug response. *Clin Pharmacol Ther* 2016; **99**: 468–70.

23 Reich DE, Cargill M, Bolk S, *et al*. Linkage disequilibrium in the human genome. *Nature* 2001; **411**: 199–204.

24 Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 1999; **286**: 487–91.

25 Evans WE, McLeod LJ. Pharmacogenomics – drug disposition, drug targets, and side effects. *N Engl J Med* 2003; **348**: 538–49.

26 Weinshilboum R. Inheritance and drug response. *N Engl J Med* 2003; **348**: 529–37.

27 Benet LZ, Kroetz DL, Sheiner LB. Pharmacokinetics. The dynamics of drug absorption, distribution, and elimination. In: Hardman JG, Limbird LE, Molinoff PB, Ruddon WR, Goodman Gilman G, eds. *Goodman & Gilman's The pharmacological Basis of Therapeutics*, 9th edn. New York: McGraw-Hill, 1996, pp. 3–27.

28 Schillevoort I, de Boer A, van der Weide J, *et al*. Antipsychotic-induced extrapyramidal syndromes and cytochrome P450 2D6 genotype: a case–control study. *Pharmacogenetics* 2002; **12**: 235–40.

29 Poulsen L, Brosen K, Arendt-Nielsen L, Gram LF, Elbaek K, Sindrup SH. Codeine and morphine in extensive and poor metabolizers of sparteine: pharmacokinetics, analgesic effect and side effects. *Eur J Clin Pharmacol* 1996; **51**: 289–95.

30 Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med* 2017; **19**: 69–76.

31 Borges S, Desta Z, Li L, *et al*. Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. *Clin Pharmacol Therapeut* 2006; **80**(1): 61–74.

32 Goetz MP, Knox SK, Suman VJ, *et al*. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res Treat* 2007; **101**: 113–21.

33 Jung JA, Lim HS. Association between CYP2D6 genotypes and the clinical outcomes of adjuvant tamoxifen for breast cancer: a meta-analysis. *Pharmacogenomics* 2014; **15**: 49–60.

34 Swen JJ, Huizinga TW, Gelderblom H, *et al*. Translating pharmacogenomics: challenges on the road to the clinic. *PLoS Med* 2007; **4**(8): e209.

35 Meyer UA. Pharmacogenetics – five decades of therapeutic lessons from genetic diversity. *Nat Rev Genet* 2004; **5**(9): 669–76.

36 McLeod HL, Krynetski EY, Relling MV, Evans WE. Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* 2000; **14**: 567–72.

37 Krynetski EY, Schuetz JD, Galpin AJ, Pui CH, Relling MV, Evans WE. A single point mutation leading to loss of catalytic activity in human thiopurine S-methyltransferase. *Proc Natl Acad Sci USA* 1995; **92**: 949–53.

38 Nguyen CM, Mendes MA, Ma JD. Thiopurine methyltransferase (TPMT) genotyping to predict myelosuppression risk. *PLoS Curr* 2011; **3**: RRN1236.

39 Lennard L. Implementation of TPMT testing. *Br J Clin Pharmacol* 2014; **77**: 704–14.

40 Hoffmeyer S, Burk O, von Richter O, *et al*. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci USA* 2000; **97**: 3473–8.

41 Siddiqui A, Kerb R, Weale ME, *et al*. Association of multidrug resistance in epilepsy with a polymorphism in the drug-transporter gene ABCB1. *N Engl J Med* 2003; **348**: 1442–8.

42 Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; **33**: 177–82.

43 Schafer M, Rujescu D, Giegling I, *et al*. Association of short-term response to haloperidol treatment with a polymorphism in the dopamine D(2) receptor gene. *Am J Psychiatry* 2001; **158**: 802–4.

44 Mukae S, Aoki S, Itoh S, Iwata T, Ueda H, Katagiri T. Bradykinin B(2) receptor gene polymorphism is associated with angiotensin-converting enzyme inhibitor-related cough. *Hypertension* 2000; **36**: 127–31.

45 Mahmoudpour SH, Veluchamy A, Siddiqui MK, *et al*. Meta-analysis of genome-wide association studies on the intolerance of angiotensin-converting enzyme inhibitors. *Pharmacogenet Genom* 2017; **27**: 112–19.

46 Cusi D, Barlassina C, Azzani T, *et al*. Polymorphisms of alpha-adducin and salt sensitivity in patients with essential hypertension. *Lancet* 1997; **349**: 1353–7.

47 van Wieren-de Wijer DB, Maitland-van der Zee AH, de Boer A, *et al*. Interaction between the Gly460Trp alpha-adducin gene variant and diuretics on the risk of myocardial infarction. *J Hypertension* 2009; **27**(1): 61–8.

48 Kurzrock R, Kantarjian HM, Druker BJ, Talpaz M. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann Intern Med* 2003; **138**: 819–30.

49 Vandenbroucke JP, Koster T, Briet E, Reitsma PH, Bertina RM, Rosendaal FR. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet* 1994; **344**: 1453–7.

50 Vandenbroucke JP, van der Meer FJ, Helmerhorst FM, Rosendaal FR. Factor V Leiden: should we screen oral contraceptive users and pregnant women? *BMJ* 1996; **313**: 1127–30.

51 Vegter S, Boersma C, Rozenbaum M, Wilffert B, Navis G, Postma MJ. Pharmacoeconomic evaluations of pharmacogenetic and genomic screening programmes: a systematic review on content and adherence to guidelines. *Pharmacoeconomics* 2008; **26**(7): 569–87.

52 Kaminsky LS, Zhang ZY. Human P450 metabolism of warfarin. *Pharmacol Therapeut* 1997; **73**: 67–74.

53 Hirsh J, Dalen JE, Anderson DR, *et al*. Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. *Chest* 1998; **114**: 445S–469S.

54 Majerus PW, Broze GJ Jr, Miletich JP, Tollefsen DM. Anticoagulant, thrombolytic, and antiplatelet drugs. In: Hardman JG, Limbird LE, Molinoff PB, Ruddon WR, Goodman Gilman G, eds. *Goodman & Gilman's The pharmacological Basis of Therapeutics*, 9th edn. New York: McGraw-Hill, 1996, pp. 1341–59.

55 Takahashi H, Kashima T, Nomoto S, *et al*. Comparisons between in-vitro and in-vivo metabolism of (S)-warfarin: catalytic activities of cDNA-expressed CYP2C9, its Leu359 variant and their mixture versus unbound clearance in patients with the corresponding CYP2C9 genotypes. *Pharmacogenetics* 1998; **8**: 365–73.

56 Steward DJ, Haining RL, Henne KR, *et al*. Genetic association between sensitivity to warfarin and expression of CYP2C9*3. *Pharmacogenetics* 1997; **7**: 361–7.

57 Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* 1999; **353**: 717–19.

58 Taube J, Halsall D, Baglin T. Influence of cytochrome P-450 CYP2C9 polymorphisms on warfarin sensitivity and risk of over-anticoagulation in patients on long-term treatment. *Blood* 2000; **96**: 1816–19.

59 Higashi MK, Veenstra DL, Kondo LM, *et al*. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* 2002; **287**: 1690–8.

60 Sanderson S, Emery J, Higgins J. CYP2C9 gene variants, drug dose, and bleeding risk in warfarin-treated patients: a HuGEnet systematic review and meta-analysis. *Genet Med* 2005; **7**: 97–104.

61 Daly AK, King BP. Pharmacogenetics of oral anticoagulants. *Pharmacogenetics* 2003; **13**: 247–52.

62 Kimmel SE, French B, Kasner SE, *et al*. A pharmacogenetic versus a clinical algorithm

for warfarin dosing. *N Engl J Med* 2013; **369**: 2283–93.

63 Pirmohamed M, Burnside G, Eriksson N, *et al*. A randomized trial of genotype–guided dosing of warfarin. *N Engl J Med* 2013; **369**: 2294–303.

64 Gage BF, Bass AR, Lin H, *et al*. Effect of genotype-guided warfarin dosing on clinical events and anticoagulation control among patients undergoing hip or knee arthroplasty: the GIFT randomized clinical trial. *JAMA* 2017; **318**: 1115–24.

65 Chanock S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis Markers* 2001; **17**: 89–98.

66 Nebert DW, Jorge-Nebert L, Vesell ES. Pharmacogenomics and "individualized drug therapy": high expectations and disappointing achievements. *Am J Pharmacogenom* 2003; **3**: 361–70.

67 Kleinbaum DG, Kupper LL, Morgenstern H. Interaction, effect modification, and synergism. In: Kleinbaum DG, Kupper LL, Morgenstern H, eds. *Epidemiologic Research Principles and Quantitive Methods*. New York: Van Nostrand Reinhold, 1982, pp. 403–18.

68 Botto LD, Khoury MJ. Facing the challenge of complex genotypes and gene–environment interaction: the basic epidemiologic units in case–control and case–only designs. In: Khoury MJ, Little J, Burke W, eds. *Human Genome Epidemiology*. New York: Oxford University Press, 2004, pp. 111–26.

69 Kraft P. Multiple comparisons in studies of gene x gene and gene x environment interaction. *Am J Hum Genet* 2004; **74**: 582–4.

70 Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001; **2**: 91–9.

71 Rao DC. Genetic dissection of complex traits: an overview. *Adv Genet* 2001; **42**: 13–34.

72 Kelada SN, Eaton DL, Wang SS, Rothman NR, Khoury MJ. Applications of human genome epidemiology to environmental health. In: Khoury MJ, Little J, Burke W, eds. *Human Genome Epidemiology*. New York: Oxford University Press, 2004, pp. 145–67.

73 Kerwin RW. A perspective on progress in pharmacogenomics. *Am J Pharmacogenom* 2003; **3**: 371–3.

74 Rebbeck TR, Ambrosone CB, Bell D, *et al*. SNP's haplotypes and cancer: applications in molecular epidemiology. *Cancer Epidemiol Biomarkers Prevent* 2004; **13**: 681–7.

75 Schreiber W, Stuehlinger HG, Brunner M, Hollenstein U. Preclinical management of thromboembolic disorders. *Semin Thromb Hemostas* 1996; **22**: 3–13.

76 Lachenbruch PA. A note on sample size computation for testing interactions. *Stat Med* 1988; **7**: 467–9.

77 Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2000; **92**: 1151–8.

78 Millikan RC. Re: Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2001; **93**: 156–8.

79 Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in case–control association studies of admixed populations. *Genet Epidemiol* 2004; **27**: 14–20.

80 Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case–control studies. *Am J Hum Genet* 2002; **71**: 304–11.

81 Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; **361**: 598–604.

82 Manolio TA, Collins FS, Cox NJ, *et al*. Finding the missing heritability of complex diseases. *Nature* 2009; **461**(7265): 747–53.

83 Thompson D, Stram D, Goldgar D, Witte JS. Haplotype tagging single nucleotide polymorphisms and association studies. *Hum Hered* 2003; **56**: 48–55.

84 Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–9.

85 Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: case–control studies with no controls! *Am J Epidemiol* 1996; **144**: 207–13.

86 Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case–only designs for assessing susceptibility in population-based case–control studies. *Stat Med* 1994; **13**: 153–62.

87 Lynch AI, Irvin MR, Boerwinkle E, *et al*. RYR3 gene polymorphisms and cardiovascular disease outcomes in the context of antihypertensive treatment. *Pharmacogenomics J* 2013; **13**: 330–4.

88 Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR; International Meta-analysis of HIV Host Genetics. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002; **156**: 204–10.

89 Capen EC, Clapp RV, Campbell WM. Competitive bidding in high risk situations. *J Petroleum Technol* 1971; **23**: 641–53.

90 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003; **100**: 9440–5.

91 Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006; **7**: 781–91.

92 Klein TE, Altman RB, Eriksson N, *et al*. Estimation of a warfarin dose with clinical and pharmacogenetic data. International Warfarin Pharmacogenetics Consortium. *N Engl J Med* 2009; **360**: 753–64.

93 Little J, Higgins JP, Ioannidis JP, *et al*. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement. *Hum Genet* 2009; **125**: 131–51.

94 Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; **65**: 220–8.

95 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.

96 Marsh S, McLeod HL. Cancer pharmacogenetics. *Br J Cancer* 2004; **90**: 8–11.

97 Gage BF, Eby D, Johnson JA, *et al*L. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Therapeut* 2008; **84**: 326–31.

98 Roses AD. Pharmacogenetics in drug discovery and development: a translational perspective. *Nat Rev Drug Discov* 2008; **7**(10): 807–17.

99 Veenstra DL. The interface between epidemiology and pharmacogenomics. In: Khoury MJ, Little J, Burke W, eds. *Human Genome Epidemiology*. New York: Oxford University Press, 2004, pp. 234–46.

100 Nuffield Council on Bioethics. www nuffieldbioethics org/home (accessed April 19, 2019).

101 Thomas DC. Genetic epidemiology with a capital "E". *Genet Epidemiol* 2000; **19**: 289–300.

102 Williams-Jones B, Corrigan OP. Rhetoric and hype: where's the "ethics" in pharmacogenomics? *Am J Pharmacogenom* 2003; **3**: 375–83.

103 Chicurel ME, Dalma-Weiszhausz DD. Microarrays in pharmacogenomics – advances and future promise. *Pharmacogenomics* 2002; **3**: 589–601.

104 Serono VP. Genomics raising disease redefinition, other orphan issues. *Pink Sheet* April 10, 2000. www.thepinksheet.com (accessed April 19, 2019).

31

Bioethical Issues in Pharmacoepidemiologic Research

Laura E. Bothwell[1], Annika Richterich[2], and Jeremy A. Greene[3]

[1] *Health Sciences Department, Worcester State University, Worcester, MA, USA*
[2] *Faculty of Arts & Social Sciences, Maastricht University, Maastricht, The Netherlands*
[3] *Department of the History of Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA*

Because the bioethical issues involved in pharmacoepidemiologic research are closely related to changing patterns of drug usage and changing technologies of surveillance and data analysis, it is impossible to understand them without attention to historical and sociological perspectives. The field of pharmacoepidemiology emerged as a result of broader recent developments in medical therapeutics, concomitant with the expansion and refinement of the field of bioethics. Some key bioethical principles relevant to pharmacoepidemiologic research have remained significant over time; others have only gained attention in recent years. This chapter briefly introduces historical and sociological dimensions of pharmacoepidemiology from an international perspective, with an eye to commonalities and differences in national variations in ethical approaches to the field.

On the most common level, it is widely believed that pharmacoepidemiologic studies should create data that benefit public health, improve drug safety, and ensure efficacy. The protection of research subjects' rights and safety, their wellbeing, dignity, autonomy and privacy, as well as the reliability and robustness of generated data are relatively universal normative cornerstones of pharmacoepidemiology ethics. The same goes for the injunction that objectives and results of pharmacoepidemiologic research should be independent from economic and promotional interests of pharmaceutical companies or device manufacturers. Yet these principles are not simple to implement systematically at an international level. In this chapter, we explore the emergence and conduct of pharmacoepidemiologic research in three major global settings in which the field developed (North America, Europe, and East Asia) and some of the key challenges, tensions, and trends in historic and current international ethical policies relating to pharmacoepidemiology.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Clinical Problems to be Addressed by Pharmacoepidemiologic Research

Emergence, Changing Methods, and Moral Stakes of Pharmacoepidemiology in 20th Century North America

In 1962, a series of epidemiological reports initiated by the German physician Widukind Lenz connected a recent increase in phocomelia, a birth defect which resulted in grossly visible limb deformities, with maternal use of the popular new antinausea medicine Contergan® (thalidomide) [1–4]. Images of thalidomide children became an international symbol of the ethical failure of the medical profession and the regulatory state to protect vulnerable populations from the harmful effects of widely marketed new drugs. Contergan had been extensively marketed to physicians and consumers alike, and its premarket testing and postmarket promotion had emphasized its remarkably *nontoxic* safety profile by available standards of clinical pharmacology [5,6]. As Lenz's work was read internationally, his careful use of the correlative techniques of infectious disease epidemiology within the terrain of prescription drug use documented not only the unseen dangers of newly marketed drugs but also the need for a new discipline of pharmaceutical epidemiology to scour observational data for therapeutic effects and adverse reactions that could clearly be associated with drug use in clinical practice [7,8].

The recognition that the risks of new drugs could be better understood when they were consumed by broad numbers of patients had been evident long before Lenz's epidemiology of thalidomide-associated phocomelia. Indeed, the history of federal drug regulation in the United States can be recounted as a succession of measures taken in response to dangers of drugs that became apparent after widespread consumption by the general public [9–11]. However, until the 1960s the Food and Drug Administration (FDA) had very limited authority in the postmarket regulation of drugs. The agency had neither direct means to control physician prescriptions nor resources to gather data on prescribing of newly marketed drugs. While the Committee on Pharmacy and Chemistry of the American Medical Association (AMA) nominally maintained more influence in both arenas, it depended entirely upon voluntary physician reports, and Committee members complained loudly that the system itself was doomed to failure; as one report noted, "physicians reported only a small fraction of all cases and the total number of patients receiving a drug was unknown" [12].

The 1962 Kefauver–Harris Amendments, passed largely on the strength of popular moral outrage over thalidomide, demanded that pharmaceutical manufacturers establish records and make reports to the FDA of "data relating to clinical experience and other data or information, received or otherwise obtained" [13] for all new drugs. By 1967, the agency had developed a protocol requiring manufacturers to seek and report any reported or published case reports related to putative side effects of their products. Any novel or unexpected adverse effect was to be reported to the agency within 15 days; other information "pertinent to the safety or effectiveness of the drug" was to be reported quarterly for the first year after approval, twice in the second year, and annually thereafter. Yet this kind of information could become actionable only after years of case reports, and then only if one of the relatively few FDA staffers took an active interest in pursuit of a specific question of drug harm.

The hospital became the center of early programs of pharmacoepidemiologic surveillance. By 1964, the FDA and AMA had built a surveillance program involving more than 600 hospitals, which became the focus of early pharmacoepidemiologic research by Johns Hopkins University's Leighton Cluff, Harvard University's Thomas

Chalmers, and Tufts University's Hershel Jick [14–18]. Yet the data were still only as good as the reporting physicians' records [19]. As Leighton Cluff noted, an early validation system of reporting efforts at the Johns Hopkins Hospital "proved completely unsatisfactory for detecting drug reactions …during recent daily intensive surveillance of one hospital service, four times as many reactions were detected than had been reported on the cards from the entire hospital" [19]. Would-be epidemiologists of adverse drug effects needed a way to circumvent the physician as reporting device and the digitization of data provided an appealing solution. Cluff's attempts at computerized drug monitoring involved the creation of three linked datasets for every drug received by every patient in a dedicated hospital ward [20]. D.J. Finney, another early theorist of computerized drug monitoring, expressed these data sets as a linked "P-D-E system," in which P(atient) population data would be systematically gathered within a set geographic or hospital catchment area, the D(rug) data would include records of all relevant prescriptions, and E(vent) collection would record all untoward reactions potentially attributable to the drugs prescribed [20].

Proponents of drug monitoring imagined a linked system of inpatient surveillance wards circling the globe, which could act as pharmacovigilance sensors, detecting early signals of possible drug harms and providing descriptive data regarding their frequency, severity, and relative strength of association. Finney predicted that surveillance would change pharmacoepidemiology from a reactive into a proactive field. Allowing that "much is due to Lenz for his discovery in 1961 [that thalidomide was associated with phocomelia]," he also boasted that "a *monitor* could have signaled a warning 1½–2 years earlier" [20]. Automated inpatient surveillance systems liberated pharmacoepidemiology from the "weak link" of the reporting physician [20]. With public and private support from the United States Public Health Service and the Pharmaceutical Manufacturers Association, Dennis Slone, Hershel Jick, and Ivan Borda demonstrated the feasibility of implementing an automated hospital-based drug monitor system in 1966 [21]. Based at the Lemuel Shattuck Hospital, the Boston Collaborative Drug Surveillance Program bypassed the physician by hiring a drug surveillance nurse "whose primary role is the acquisition of accurate data" [21,22]. The Boston team became a model for an automated drug surveillance program that functioned "largely independent of clinical judgment in establishing a connection between a drug and an adverse event" [23].

Early results showed that drug-related events were both more frequent and less severe than had previously been anticipated. More than one-third of patients on the Shattuck wards experienced at least one drug-associated adverse reaction during the first year of study [24]. By 1967, the Boston group had established a numerator/denominator approach for comparing drug usage between long-term and acute hospitals through a network of five hospitals in Boston [25]. By 1968, over 2500 patients had been entered and discharged from the surveillance system, with over 26 000 monitored drug exposures, representing more than 700 individual drugs [22]. Commonly prescribed drugs, such as digoxin and heparin, could be reported in detail, yielding novel information related to their clinical pharmacology and their interactions with other drugs [26–28]. The system enabled the observation of not only obvious drug reactions (such as a rash) but also other clinical events (such as heart attacks or kidney failure) that could only be associated with drugs by careful epidemiologic surveillance.

As the Boston Collaborative Drug Surveillance Program escalated its activities and exported its methods to other sites, these new data provoked a series of drug scandals that emphasized both the utility and the limitations of

the new forms of pharmacovigilance. Clioquinol, an antiinfective that had been in use since the 1930s, was found to be associated with subacute myelooptic neuropathy in 1970, over three decades after its initial introduction. An association between the synthetic estrogen diethylstilbestrol (DES) and a rare form of cervical clear cell adenoma was reported in 1971, with evidence of a 20-year latency period between use of the drug and detection of the cancer [29]. The beta-blocker practolol became the focus of a scandal after it was associated with a potentially fatal inflammation of the skin and soft tissues (oculomucocutaneous syndrome) some five years after its broad release on the British market. These examples simultaneously elucidated the scientific and ethical necessity for drug surveillance units and underscored the impossibility of inpatient surveillance systems to capture drug–disease associations in which three decades or more might pass between drug exposure and adverse events. As Jick warned, in a systematic proposal for the theory and design of the emerging field of pharmacoepidemiology, the ability to study "drug–illness relations" required distinct methods depending on the time course and prevalence of prescription-related adverse events. High-frequency events in high-prevalence diseases could be detected swiftly by case report, low-frequency events in high-prevalence diseases required careful active ongoing surveillance, and low-frequency events in low-prevalence diseases might simply never be adequately described [23]. Many early pharmacoepidemiologic researchers viewed scientific quality and ethics as complementary: more rigorous data collection of drug-related events carried ethical benefits by enhancing medical practitioners' capacity to "do no harm" to patients. As early pharmacoepidemiologic work also coincided with the development of bioethics as a field, critical principles of informed consent, external review of research protocols, and protection of patient privacy began to influence pharmacoepidemiologic investigators' thinking in the US and internationally.

To address the growing problems of drug safety, prescription surveillance needed to extend outwards: spatially, from the monitored wards of the hospital to the messier universe of outpatient care; temporally, from links visible in days or weeks of measurable hospital time to the longer stretches of months and years required to understand the impacts of chronic medication use; and thematically, from the isolated connection of drug and disease to the study of all steps of diagnosis, prescription, adherence, consumption, and presentation that might extend in between. In the United States, this project would find its boldest form in the Joint Commission on Prescription Drug Use, formed in response to a press conference held by Senator Edward Kennedy in November 1976, at which he announced that the new science of drug utilization studies had provided irrefutable evidence that prescription drugs were ill-used in American society [30]. Kennedy called for Congress to work with the medical profession and the pharmaceutical industry to sponsor a public–private body of expertise whose explicit purpose would be to establish a postmarket surveillance system for prescription drugs [31]. As the Commission would note in its final report, the purpose of systematic prescription surveillance was "not merely to learn 'something' about a drug but to glean information that is useful in improving the rational use of drugs" [31].

Conceived as a public–private venture, the Commission ran from 1976 until 1979 and issued its final report in the first month of 1980. The Commission worked to integrate the social, epidemiological, marketing, and policy interests in prescriptions as a source of data. Initially, the prospects for a harmonization of these four perspectives seemed auspicious. At the first meeting, Howard L. Binkley, Vice President for Research and Planning of the Pharmaceutical

Manufacturers Association, provided a description and critique of presently available sources of data on trends in the prescribing and dispensing of prescription drugs, with an emphasis on how market research data could be linked to broader systems of private and public claims and outcomes data [31]. Yet as the Commission assessed its findings by 1979, it became clear that although several datasets existed, no individual dataset contained enough information to deliver sufficient granularity to allow the full assessment of drug use in outpatient practice.

The Commission began to interview hybrid data sources that illustrated new links between the public and private nature of prescriber data sets. Fledgling HMOs such as Kaiser Permanente and the Group Health Cooperative of Puget Sound developed in-house proprietary databases that linked both prescription claims and outcomes data in the same place [31]. Exploratory work by Hershel Jick following the use of the blockbuster antiulcer drug Tagamet® (cimetidine) in Puget Sound pharmacies suggested that this approach could be quite promising [32]. Another hybrid form was introduced by Noel Munson, a spokesman from Prescription Card Services (PCS), a private prescription data company that acted as a "fiscal intermediary" for public payment groups like Medicare and Medicaid and other groups that paid for prescription drugs. But these individual companies (e.g., PCS) appeared to code their data according to their own proprietary software [31]. Even within the Medicaid system, the promise of effortless data linkage remained a dream in the late 1970s, complicated by wide state-by-state discrepancies in patterns of coding, storing, and retrieving prescription data [31].

If the 1980 publication of the Joint Commission report represented a high point of collaboration between market researchers, epidemiologists, policy reformers, and sociologists in imagining an early "big data" universe for therapeutic surveillance, it also represented a dream of collaborative work that would soon dissipate. Like many other grand designs for federally sponsored health programs conceived in the later 1970s and proposed in the early 1980s, its speculative structures would never materialize, its measures would be left unfunded, and subsequent calls for a center for postmarketing surveillance would be repeated, and unfunded, every few years for several decades. Only in the past decade, with the passage of the Food and Drug Administration Amendments Act of 2007 (FDAAA), would a substantial US public investment be made in the construction of a linked public prescription database for pharmacoepidemiologic research, with the creation of the FDA's new automated pharmacovigilance program, the Sentinel Initiative, which officially launched in 2016 (see Chapter 25).

European Pharmacoepidemiologic Trends and Ethics

In Europe, several nations with centralized national health systems like England and Sweden created prescription surveillance systems by the second half of the 20th century. Scandinavian countries in particular had long histories of centrally organized pharmacy records and more tightly controlled national formularies of allowable drugs [33]. Moreover, the World Health Organization had set up a regional European Drug Utilization Group in Oslo which held a prominent conference on the overprescribing of prescription drugs in 1969 [34] and then proceeded to develop methods of comparing utilization across drug classes and across national pharmacy standards [35]. Ironically, even in countries such as Sweden, much of the prescription data came from the private sector [33,36,37]. Still, pharmacoepidemiologic research in Europe continued to receive substantial public support throughout the 1970s, 1980s, and 1990s.

The founding of the European Medicines Agency (EMA) in 1995 was a crucial step toward a pan-European supervision of medicines. The

decentralized agency is critical to the European Medicines Regulatory Network (EMRN), partnering with the European Commission (EC) and national authorities of European Economic Area (EEA) member states (the Heads of Medicines Agencies [HMA] network). The EMRN's main objective is to achieve a consistent approach to medicines regulation across the EU. In collaboration with network partners, the EMA oversees the scientific evaluation, safety and efficacy monitoring of human (and veterinary) medicines in the EU. For most innovative medicines, including those for rare diseases, a central assessment and marketing authorization coordinated by the EMA is compulsory. In cases of human medicines, the EMA's Committee for Medicinal Products for Human Use (CHMP) carries out a scientific assessment, based on which the EC decides whether to grant marketing authorization. Once granted, such a centralized marketing authorization is valid across the EU. Predominantly, though, medicines in the EU are authorized by member states' national authorities.

Shared, key ethical requirements in European pharmacoepidemiologic research came to include beneficence, transparency, scientific independence, and integrity. Yet, inconsistent application and authorization procedures for clinical studies in European Union (EU) and European Economic Area (EEA) member states have long been criticized. This also applies to pharmacoepidemiology and pharmacovigilance. Especially for multinational, noninterventional studies (NIS), it has been lamented that "… a patchwork of regulations and codes of conduct have to be followed" [38].

Partly in response to some of these issues, since the early 2000s new EU regulations, directives, and guidelines have been introduced. These aim to facilitate ethical, effective pharmacoepidemiologic practices in and across different member states. Currently, crucial regulatory changes are under way that will affect pharmacoepidemiology and pharmacovigilance in the EU.

The EU pharmacovigilance legislation aims to minimize risks and harms posed by adverse drug reactions (ADRs). Its implementation is overseen by the EMA, EU member state authorities, and the European Commission (EC). Key legal documents for the pharmacovigilance legislation and pharmacoepidemiologic studies are EU Regulation No 1235/2010 and Directive 2010/84/EC [39]. In effect, the regulation outlines measures for safeguarding patients' safety and rights and asserts the crucial role of healthcare professionals in reporting ADRs. It moreover acknowledges the necessity to develop EU/EEA-wide "… harmonized guiding principles for, and regulatory supervision of, postauthorization safety studies that are requested by competent authorities and that are noninterventional, that are initiated, managed or financed by the marketing authorization holder" [39]. Among other deliverables, the regulation established the EudraVigilance database as a main platform for the obligatory reporting of ADRs by marketing authorization holders and respective national authorities.

In response to the benfluorex scandal, the legislation was amended in 2012. Servier Pharmaceuticals' Mediator® (benfluorex), marketed as an add-on for diabetes and hyperlipidemia, was under pharmacovigilance investigation in France since 1998. It was found that the drug caused cardiovascular complications in 2003. In response, Servier did not reapply for marketing authorization in Spain and Italy, effectively withdrawing the product from the market in those countries. However, benfluorex continued to be available and approved for diabetes treatment in France and other countries until 2009, when its authorization was fully revoked; its efficacy was found to be limited and it risked causing cardiac valvulopathy [40]. Subsequently, EU Regulation No 1027/2012 and Directive 2012/26/EC were published, amending the 2010 EU pharmacovigilance legislation. The amendments especially addressed the issue that safety measures for medicinal products

need to be implemented consistently and in a timely fashion in all member states where respective products were authorized.

The benfluorex scandal points to broader challenges regarding pharmacovigilance and pharmacoepidemiologic research in the EU: regulations and guidelines need to be applied across multiple states and to different actors, including national marketing authorization holders and applicants. While the legislation outlines fairly broad objectives, responsibilities, and issues, these are specified in concrete deliverables. One of these deliverables was the founding of the EMA Pharmacovigilance Risk Assessment Committee (PRAC) which monitors and assesses drug safety in the EU. Moreover, it initiated the development of the EMA's Good Pharmacovigilance Practices (GVP) guideline (described later in this chapter).

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) was established in 2006 and is coordinated by the EMA. It is an expertise and resource network focused on pharmacoepidemiology and pharmacovigilance in Europe. It consists of partners that are public and not-for-profit organizations, including research and pharmacovigilance centers, university hospitals, healthcare database hosts, and electronic registry sponsors. For-profit organizations, such as contract research institutions, may only participate if they conduct pharmacoepidemiologic and/or pharmacovigilance studies commissioned by third parties. While pharmaceutical companies are not eligible for becoming ENCePP partners, the network provides relevant resources and allows for these companies to be involved in public document reviews.

The ENCePP offers crucial guideline documents for pharmacoepidemiology and pharmacovigilance: a Code of Conduct; the ENCePP Checklist for Study Protocols; and the ENCePP Guide on Methodological Standards in Pharmacoepidemiology. The Code lays down rules and principles aimed at ensuring transparency and scientific independence. While adherence to the Code is voluntary, it is required to receive the ENCePP Seal. Conditions for receiving the Seal are, among others, that a study is entered in the EU PAS Register and that it is of scientific and public health relevance, rather than mainly pursuing results which may promote certain medicinal products. The Checklist is meant to ensure that studies adhere to epidemiological principles, while also considering methodological transparency and the need for public outreach.

East Asian Pharmacoepidemiologic Trends and Ethics

East Asia has made major contributions as the field of pharmacoepidemiology has grown, producing a very robust body of pharmacoepidemiologic research that has expanded in recent decades. Researchers in South Korea, Japan, and Taiwan have linked into comprehensive data systems on insurance claims created through universal insurance coverage of these entire national populations. To help protect patient privacy, these databases have been made available for drug safety research only to researchers in nonprofit organizations who must apply and undergo ethical review [41].

The Korea Food and Drug Administration (KFDA) launched an adverse drug reaction (ADR) reporting system in 1988, although the reporting rate was initially very low. In 2004, the KFDA mandated that pharmacists and pharmaceutical companies report adverse drug reactions. The KFDA also established regional pharmacovigilance centers in university hospitals that now provide nearly complete coverage of the country. The KFDA funded a pharmacovigilance research network (PVNet) among these centers, and researchers in the network use their data for studying adverse events. The Korean national health insurance database also contains all information on insurance claims

made and prescriptions for approximately 50 million Koreans, and this has been used for pharmacovigilance [41].

In Japan, drug manufacturers are required to report adverse drug reactions to the Pharmaceuticals and Medical Devices Agency (PMDA). A partial adverse drug reaction dataset is available to researchers through the PMDA website. Healthcare professionals report adverse drug events to the Ministry of Health, Labor and Welfare. Japan made its national insurance claims database available for drug safety researchers in 2011. The database covers the entire population of 128 million and includes basic patient characteristics, drug prescription and dispensing, medical procedures, hospital admission, and annual health check data (for some patients) [41]. To protect patient privacy, Japan's national database is usually not available for purchase and may only be shared in some cooperative research projects [42]. The Japanese government has also created the Medical Information for Risk Assessment Initiative (MIHARI) to access data from different sources and create a central database with a common data format [43].

Taiwan requires mandatory reporting of serious adverse reactions by medical institutions, pharmacies, and drug and device companies, as well as obligatory safety reports for newly marketed drugs over a five-year surveillance period. In Taiwan, the National Adverse Drug Reactions Reporting System has been the primary source for postmarketing surveillance of adverse drug events. Taiwan's single-payer National Health Insurance (NHI) program was created in 1995 and covers more than 99% of the population. The NHI Research Database is thus a highly comprehensive dataset including basic patient data, care record and expenditure claims, and pharmaceutical reimbursements. There are also subject datasets available to researchers on topics such as traditional Chinese medicine, cancer, diabetes, dental, catastrophic illness, and psychiatric care. Patients and medical facilities are deidentified for pharmacoepidemiologic research use of the NHI Research Database [41]. To protect patient privacy, researchers using Taiwan's NHI Research Database also receive data for 10% or less of the population. Ethical policies for data privacy stipulate that no individual-level data can be shared with researchers from other countries [42].

China and other East Asian countries also have been creating national healthcare claims databases [44]. In China, the Shanghai Center for Adverse Drug Reaction Monitoring has operated a drug surveillance and evaluation system since 2001 that works with patient information from 10 Shanghai hospitals [43]. The Asian Pharmacoepidemiology Network (AsPEN) was recently established as a multinational research network for pharmacoepidemiological research that promotes international communication among academia, government, industry, and consumers. The network functions to promptly identify drug safety issues [44].

Pharmacoepidemiology ethics in East Asia are similar in many ways to those of Western countries, including features such as institutional ethical review and guiding principles such as beneficence, justice, autonomy, and data privacy. However, experts on East Asian bioethics also have recognized some distinctions. For example, scholars have contended that much East Asian bioethical thinking reflects value systems that emphasize the family and public interest ahead of the individual rights of the liberal subject that characterize much of Western bioethics. The family is often depicted as responsible for taking care of members who become sick, and medical decision making has often been family based. Some also have noted a plurality of ethical perspectives within East Asia, contending that a simple Eastern and Western bioethical dichotomy of communitarian versus individualistic values would be overly simplistic. Others have viewed bioethics as a Western entity, promoting the development of

Asian bioethics based more on the traditions, philosophies, religions, and perspectives of the region's cultures [45]. Future policies should consider these issues as core principles for pharmacoepidemiologic research ethics are discussed.

Methodologic Problems to be Solved by Pharmacoepidemiologic Research

More work remains to establish international ethical policy harmonization while also promoting practices that support cultural variation in ethical values. Yet, as pharmacoepidemiological practices developed in different national contexts that have been incorporated into increasingly globalized flows of pharmaceuticals and pharmaceutical-related information, a number of ethical principles and practices have been adopted widely across international settings in efforts to establish consistent pharmacoepidemiologic methodology.

The expansion of the field of pharmacoepidemiology has coincided with the establishment and institutionalization of the discipline of bioethics. Numerous critical ethical concepts took hold early in pharmacoepidemiology and have remained significant over time. For example, privacy of medical data is a historically consistent value guiding the ethics of global pharmacoepidemiologic research. Pharmacoepidemiologic research protocols and/or database designs also often have been subjected to review by institutional review boards as external review has become increasingly widespread for biomedical research since the second half of the 20th century, although there is variation in the nature of this review. For example, some pharmacoepidemiologic research has been reviewed by institutional or national ethics boards, as well as by privacy boards [46]. Some countries also do not require ethical review for deidentified datasets [47].

Informed Consent

Informed consent became increasingly valued as a critical standard of international research ethics following its establishment as a cornerstone of the 1964 Declaration of Helsinki, a ground-breaking statement of international human experimentation ethics [48]. However, the role of informed consent has been perceived differently in interventional versus noninterventional research studies. Many ethicists of international human subject research have argued that since pharmacoepidemiologic research involves relatively low risks to participants, patient consent is necessary only for studies that involve contact with patients/research subjects, such as for direct intervention or prospective gathering of information. There has been a broad acceptance among ethicists allowing researcher access to identifiable medical records for pharmacoepidemiologic research without explicit individual subject authorization [46]. Research has also found that public opinion has echoed the views of professional ethicists that pharmacoepidemiologists should be permitted to use identifiable patient records, without patient consent, to study drug safety as long as existing ethical guidelines and relevant laws are followed [49].

A number of nations, however, require explicit informed consent from each study participant, and there are also international variations in requirements for electronic consent versus hard copy written consent. Ethical regulatory disharmony causes differences in study conduct between countries and increases the cost of assembling multinational data. This poses challenges for conducting large international studies capable of detecting rare events. Additionally, requirements of explicit individual informed consent are problematic in that they can corrupt data by preventing a postmarketing pharmacoepidemiologic study from detecting fatal or serious events since people who have died are unable to provide informed consent [47].

Thus, it is unsurprising that ethicists weighing risks and benefits have tended to contend that individual consent is not essential for use of patient records in pharmacoepidemiologic research.

However, over time it has become normative that pharmacoepidemiologists also must meet certain requirements when conducting research in which participant consent is waived. These requirements often include that the use of protected health information involves no more than minimal risk to patients, the research could not be effectively conducted without access to the protected health information and/or the waiver of individual consent, the privacy risks to individuals are reasonable in relation to any value to the individuals of the knowledge expected to result from the study, there is a sound plan to protect patients from improper use or disclosure of their information, there is a plan to destroy identifiers at the earliest opportunity consistent with the research, and the data will not be shared with external parties to the research [46].

Recent attention has been given to waiver of patient informed consent to use data on substances of abuse or drugs that carry social stigma. Patient privacy is essential in these areas of research; however, requiring informed consent for each patient or allowing retraction of sensitive drug information from patient records leads to partial datasets that impede the ability of researchers to study the impact of these substances on patient health outcomes. The negative consequences of failing to collect sound pharmacoepidemiologic data on the health effects of these substances are likely worse than the relatively minimal risk associated with waiver of patient consent. However, in such circumstances, the highest precautions should be taken to protect patient privacy, such as deidentifying data through secure codes or potentially having extra ethics training requirements for all researchers using data on stigmatized or abused substances. (See Chapter 28 for further discussion of pharmacoepidemiology research on drugs of abuse.)

Ethics of Surveillance

Surveillance has long provoked public concern regarding privacy, confidentiality, and autonomy. This is relevant to postmarketing surveillance, since health information is seen as highly sensitive and personal. Thus, pharmacoepidemiologic researchers need to balance possible risks to a larger population against the harms concerning individuals, such as a possible infringement of privacy. While privacy is highly important to the ethics of pharmacoepidemiologic research, privacy is not an absolute value, nor does it seem to have been perceived as such in public health surveillance history. Rather, privacy is one of multiple values that are balanced in public health surveillance [50]. It has been argued that ensuring privacy is part of the broader value of protecting autonomy. Yet other key principles to be balanced in pharmacoepidemiologic research include beneficence to promote research that adds to the existing knowledge base of medicine to improve patient health and prevent mortality; nonmaleficence, or the prevention of patient harm; and justice, which manifests as the fair distribution of research burdens and benefits among people [51].

Risks of surveillance can be minimized through confidentiality and data anonymization. Such strategies are ethically imperative, since they safeguard individuals' rights, privacy, autonomy, and dignity. Applying the highest ethical standards and communicating with the public about potential criticism are also important for a positive public perception of pharmacoepidemiology.

While there have been some disagreements, international ethics policies have developed some common stances toward ethical review of drug surveillance. Certain pharmacoepidemiologic research tends to qualify as exempt from ethics board review or qualifies for expedited review by an ethics board chair or a designated member. For studies in which it is not possible for investigators to identify individual patients, ethics board review is often not required. For example, the US 45 Code of Federal Regulations 46.101

exempts from institutional review "research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects" [52]. In many countries, research is also often eligible for expedited review if it poses no more than minimal risk to patients and involves a retrospective analysis of existing records. Still, ethics review policies vary internationally and by institutional practice, depending *inter alia* on respective national/state regulations, posing challenges for global collaborative studies [53]. This may lead to inconsistent risk–benefit assessments and variations in balancing subjects' protection (e.g., regarding safety and privacy) against public health interests.

The European Medicines Agency's guideline on Good Pharmacovigilance Practices (GVP) provides a useful differentiation between "active surveillance" and "passive surveillance." Active surveillance is defined as a continuous, systematic process of monitoring adverse events in a population. For example, a risk management system may be put in place which allows for the active surveillance of patients receiving a medicinal product. Another active surveillance option would be the monitoring of laboratory reports to detect adverse events. Active surveillance may be part of interventional or noninterventional studies (NIS). Passive surveillance, based on patients' spontaneous reporting, for example, is commonly seen as less effective, because it runs the risk of delivering less comprehensive data [54].

Ethical Benefits of Pharmacoepidemiologic Research for Data Integrity

From a broader ethical perspective, it is increasingly clear that the expansion of pharmacoepidemiological research can provide added benefits to drug research by detecting groups at risk for adverse events. Thus, the field can play an important role in reducing drug safety data inequalities. For example, expanding drug outcomes data for groups such as minorities or small/rare genetic subpopulations who may have treatment outcome variations that can only be identified and/or adequately quantified and measured through large postmarketing pharmacoepidemiologic studies may provide substantial benefits for members of these populations. There are also limited data on the efficacy and safety of drugs in children due to the fact that historically, children have often not been included in randomized controlled trials (RCTs). Pharmacoepidemiologic research helps to fill these research gaps [55]. However, it would be ethically problematic for pharmacoepidemiology to be relied on solely to provide missing data on children, minorities, or other subgroups in lieu of RCTs, particularly in cases when RCTs could produce more robust data.

Further, pharmacoepidemiologic studies are usually conducted after drug approval, and there is high variability in the frequency and design of postmarketing pharmacoepidemiologic research [56]. Such studies are not necessarily required, and so are not a consistently reliable source of information on drug outcomes among diverse demographic groups. Clinical trials are usually required for drug approval and are thus a mechanism for ensuring broader implementation of policies requiring the inclusion of diverse research subjects [48]. Ultimately, consistent with recurring concerns over ethical practices in pharmacoepidemiologic research in general, ethicists have noted that pharmacoepidemiology related to subpopulations would benefit from a more explicit legal ethical framework, particularly to clarify ethical requirements for data sharing [57].

Problems of Conflicts of Interest for Drug Industry Research

Academia–industry collaborations have become a critical area of concern for the ethics of pharmacoepidemiologic research, particularly in

recent decades as pharmaceutical profits have soared and the stakes have been raised for the outcomes of research on drug safety and efficacy. There is an inherent conflict of interest in research that is funded by drug companies to assess their own products. Academic settings in which researcher success and advancement depend on obtaining external funding also can exacerbate the ethical problems resulting from direct relationships between drug companies and the pharmacoepidemiologists evaluating their products. Investigators in such environments are under professional pressure to secure funding, and in a climate of heightened competition for public funding sources, an academician who establishes a positive working relationship with a pharmaceutical research sponsor may increase his/her chances of obtaining future funding from that sponsor.

This creates an incentive, whether subconscious or acknowledged, for researchers to conduct studies that sponsoring drug companies will find favorable. Indeed, studies have shown a trend toward more favorable efficacy results and conclusions for industry-sponsored drug research than research sponsored by other sources, finding a bias in industry-funded research that cannot be otherwise explained by standard assessments of risk of bias [58,59]. There are a number of feasible solutions to address the ethical conflicts of interest in industry-funded research.

Currently Available Solutions

Good Pharmacoepidemiology and Pharmacovigilance Practices

The International Society for Pharmacoepidemiology (ISPE) has created Guidelines for Good Pharmacoepidemiology Practice (GPP), which provide a model for key pharmacoepidemiologic research ethics policies. The guidelines recommend that researchers include a description of quality control procedures; plans for protecting human subjects; confidentiality provisions; ethical conditions under which a study would terminate; the use of Data Safety Monitoring Boards where appropriate; institutional review board and informed consent considerations in accordance with local laws; research study registration; and plans for disseminating study results [60]. However, ISPE GPP policies are nonbinding and therefore do not resolve concerns regarding national variations in ethical oversight and requirements by regulatory agencies for postmarketing pharmacoepidemiologic work [47].

European Union policies provide a useful example of transnational efforts at regulatory standardization of good pharmacovigilance practices. EU documents concerning biomedical research in general and pharmacoepidemiologic research commonly speak of two types of clinical studies, broadly speaking: interventional, that is, experimental, and noninterventional, sometimes called observational research. On the one hand, pharmacoepidemiologic research relies on noninterventional study designs such as case–control or cohort studies. On the other hand, interventional RCTs are an important element of postmarketing pharmacoepidemiology studies (see Chapter 32).

The EMA defines Good Pharmacovigilance Practices (GVP) as "a set of measures drawn up to facilitate the performance of the safety monitoring of medicines in the European Union" [54]. It includes chapters on pharmacovigilance processes as well as product- and population-specific considerations. For EU pharmacoepidemiologic postauthorization safety studies (PASS), module VIII is particularly relevant. PASS may be interventional or noninterventional. Although the module touches upon interventional studies too, emphasis is put on noninterventional PASS.

In accordance with the EU pharmacovigilance legislation, the GVP stipulates that the EMA needs to ensure that protocols and abstracts of PASS results are published. While the primary/

lead investigator is responsible for the information provided, the registration may be made by, for example, research center staff or representatives of pharmaceutical companies funding a study. Where possible, this should be done before the study commences. Practically, registration and publication are processed through the EU postauthorization study (PAS) register, hosted by the ENCePP [61]. As the ethics review procedure and requirements for respective committees depend on national legislation, information on individual application procedures is not included in the GVP. While there is no EU regulation or directive for NIS, interventional studies are covered in the Clinical Trials Regulation.

In the European Union, methodological, ethical, and legal requirements for pharmacoepidemiologic research hinge significantly on whether a study is categorized as a "clinical trial" or as "noninterventional/nonexperimental." Both categories are defined as "clinical studies" aimed at discovering or confirming the (adverse) effects of medicinal products [62]. For pharmacoepidemiologic studies involving clinical trials, the introduction of the EU Clinical Trials Regulation (CTR) No 536/2014 will be decisive [63].

The CTR was adopted on 16 April 2014 and entered into force on 16 June 2014. According to the EMA, it will come into application in late 2019, starting a transition period of three years [64]. It is meant to harmonize research practices and to ensure the highest methodological and ethical standards across all EU as well as EEA EFTA member states. To what extent it will deliver on these promises is under discussion [65,66]. The regulation replaces the Clinical Trials Directive 2001/20/EC which is said to have "… failed to achieve its goal of simplifying the scientific and ethical review of clinical trials in the EU" [67].

Moreover, the ENCePP had problematized the NIS definition given in the 2001 directive. The ENCePP raised the issue that the definition was not sufficiently specific and created uncertainty as to what counts as NIS or RCT. Pharmacoepidemiologic prospective case–control studies – like the IPPHS investigation of primary pulmonary hypertension (PPH) occurrence in association with anorectic agents – would classify as a clinical trial according to the 2001 directive. Its ambiguous NIS definition was thus criticized for impeding the conduct of pharmacoepidemiologic studies [68].

The ENCePP Guide on Methodological Standards in Pharmacoepidemiology (Revision 6, July 2017) lays down rules and principles for transparency and scientific independence. Chapter 9 of the Guide deals with ethical aspects of pharmacoepidemiology, focusing on patient and data protection (9.1) and scientific integrity and ethical conduct (9.2). It identifies key values based on documents such as the ADVANCE Code of Conduct for Collaborative Vaccine Studies, the GPP of the International Society for Pharmacoepidemiology, and the Good Epidemiology Practice (GEP) guidelines of the International Epidemiological Association. The Guide highlights that "principles of scientific integrity and ethical conduct are paramount in any medical research" and points out that the above-mentioned ENCePP code of conduct "… offers standards for scientific independence and transparency of research in pharmacoepidemiology and pharmacovigilance" [69]. In addition, it highlights core values, such as best science, strengthening public health, and improving transparency, as stressed by the ADVANCE Code of Conduct. It also emphasizes the need for ensuring scientific autonomy, beneficence, nonmaleficence and justice, according to the four general ethical principles defined in the GEP guidelines.

Protections Against Conflicts of Interest for Drug Industry-Sponsored Research

While industry-sponsored research creates real challenges for conflicts of interest, industry also has an interest in maintaining public trust in product integrity, as well as in compliance with

regulatory ethical and methodological requirements to obtain drug approval. Thus, there is some incentive for industry to address concerns about conflicts of interest. The Board of Directors of the International Society for Pharmacoepidemiology has published a set of principles for academia–industry collaboration that can be helpful in managing industry conflicts of interest. It includes the importance of transparent research agreements, open and complete disclosure of conflicts of interest, registration of research protocols in public sites such as the ENCePP registry or ClinicalTrials.gov, compliance with local laws, clarity on confidentiality of proprietary information while also ensuring reporting of all relevant and important information to regulators, the potential value of having a steering committee and/or an independent advisory committee to the research, and an obligation to disseminate and publish research findings of potential scientific or public health importance irrespective of results [70].

While all these principles are helpful in managing financial conflicts of interest, they do not eliminate the inherent problem of drug companies having a stake in the outcomes of research that they sponsor or the ethical concerns associated with the power dynamics of industry directly funding investigators as described above. To eliminate these underlying ethical problems, the direct relationships in which companies fund individual investigators to assess specific products would need to be severed. Alternative models that eliminate these ethical conflicts can be easily envisaged. For example, the British Drug Safety Research Unit (DSRU), an independent charity supported by the National Health Service, conducts publicly funded pharmacoepidemiologic research [43]. Still, the organization conducts a large amount of research funded by unconditional donations from pharmaceutical companies. However, the companies have no control on the conduct or the publication of studies conducted by the DSRU [71] which helps to mitigate the pressure of inherent conflicts of interest in industry-funded research.

Given that industry funding may lead to biased study results, a comprehensive solution could build from the DSRU model, for example by requiring sponsors of new drugs to contribute an unconditional fee to drug regulators that would fund pharmacoepidemiologic research. By making such contributions mandatory rather than voluntary, investigators could conduct studies without concern as to whether results may influence future industry donation decisions. In the US, for example, the expansion of the FDA's Prescription Drug User Fee could easily establish a fund for pharmacoepidemiologic research.

The Future

The ethical conduct of pharmacoepidemiologic studies is of crucial importance for subjects' safety, health, and wellbeing. Moreover, it is decisive for the public perception of pharmacoepidemiology. Research in this field is rooted in the moral obligation to preempt or at least minimize medicine-related harms and health hazards. Implementing highest ethical standards helps to avoid potential damage to the public image of the field and public trust in claims of pharmacoepidemiological research as a disinterested form of expert knowledge. Such damage may be related to research practices compromised by economic interests or misconduct of the pharmaceutical industry. Thus, scientific integrity, independence, and transparency will continue to be crucial for the ethics of pharmacoepidemiologic research.

Even in the recent past, regulatory amendments relevant to pharmacoepidemiology and pharmacovigilance were often triggered by scandals, although a dream to make pharmacoepidemiology a proactive rather than a reactive field can be traced back to the 1960s if not earlier. Adjusted, new, and emerging regulations and guidelines aim at promoting ethical

pharmacoepidemiologic research that effectively identifies and reports ADRs, thus allowing for timely responses. New policies must also be more thoroughly transnational and attentive to global variations in ethical beliefs. A main challenge is and will continue to be to translate inevitably general documents into practical instructions and relevant local practices.

In the future, national regulatory authorities, universities, and research centers will continue working to align requirements towards coherent pharmacoepidemiologic research ethics. It is to be expected that further regulatory efforts will be invested in streamlining requirements for ethics review boards and ethical guidelines for noninterventional studies, especially across the EU. Although recent regulations and directives in the EU hope to address several pressing issues, many of these are complicated anew by the UK's announced withdrawal from the EU. This has already triggered practical changes, such as the relocation of the European Medicines Agency from London to Amsterdam in March 2019. Moreover, legal uncertainties are increasing [72], as it has been disclosed by the UK Department for Exiting the European Union that the post-Brexit guidelines for clinical studies in the UK may deviate from EU legislation [73].

Transparency has been stressed as a key element for ensuring ethical pharmacoepidemiologic practices. Moreover, data sharing is pivotal for effective pharmacoepidemiology and pharmacovigilance. At the same time, researchers are required to safeguard subjects' privacy and dignity. Developments such as the open data movement on the one hand and regulations aimed at protecting individuals' privacy on the other hand put researchers in a difficult position. At an increasing rate, there is a tendency to require public accessibility of scientific results and even data. Simultaneously, privacy concerns and potential regulations may pose challenges for data (re-)use in pharmacoepidemiologic studies [74].

Heightened attention has already been paid to the environmental, polluting effects of pharmaceutical residues. Regulatory documents, such as the EU pharmacovigilance legislation, acknowledge that "the pollution of waters and soils with pharmaceutical residues is an emerging environmental problem" [39]. Research examining the adverse effects of pharmaceuticals on the environment has been labeled *pharmacoenvironmentology*. With its focus on the environmental impact of drugs given at therapeutic doses, it is considered part of pharmacovigilance [75]. Assuming that environmental issues will continue to be high on the political and scientific agendas, pharmacoepidemiologic expertise will be increasingly needed to assess medicines as pollutants. In this context, pharmacoepidemiologists will need to employ and expand their methodological repertoire for studies investigating medicines' adverse effects on the environment. This development might also imply an amplified need for novel, interdisciplinary research collaboration involving pharmacoepidemiologists.

Such collaboration is also characteristic for another emerging intersection, between pharmacoepidemiology, computer, and data science. Research at the intersection of digital services, big data, and public health is a potentially promising but precarious field. It has been demonstrated that emerging digital data sources like social networking sites can function as complementary resources for pharmacoepidemiology. The use of such data sources, often referred to as a type of big data, is atypical for pharmacoepidemiologic studies but may become more common in the future. Research drawing on big data may take place outside medical departments or hospitals, for example being conducted by data scientists. Big data and emerging data science approaches have created new possibilities for pharmacoepidemiologic research. For example, Freifeld *et al*. used data from the social networking site and microblogging service Twitter to monitor ADRs [76].

The term *big data* has become associated with various leaks and scandals. The UK Science and Technology Committee concluded in a 2015 report that data misuses and leaks have led to public skepticism concerning the use of big data [77]. Not only such negative connotations, but also scientific concerns regarding users' consent, autonomy, and privacy raise ethical questions about big data research. Pharmacoepidemiologic research involving big data requires careful ethical considerations for the individuals generating such data, for example users of social networking sites. Moreover, pharmacoepidemiologists need to consider the biases inherent to digital data sources: such bias can be caused by big data retrieved from populations that do not allow for generalizations. For instance, since individuals included in a digital data sample may represent only those using an expensive/innovative technical device or service, these users could be on average younger or above average in access to health-promoting resources [78]. In addition, the quality of such data may differ from other sources of data (e.g., medical records).

Research involving these alternative sources of data is subject to different laws and regulatory frameworks when conducted in different global settings. For the US, access to health-relevant information via social networking sites such as Facebook is at present legally possible, due to the lack of protection for health-relevant data retrieved outside the traditional healthcare and research system. With regard to medical privacy, the Electronic Frontier Foundation (EFF) points out that social networking sites and other online services compromise US citizens' control over their health data:

> The baseline law for health information is the Health Insurance Portability and Accountability Act (HIPAA). HIPAA offers some rights to patients, but it is severely limited because it only applies to an entity if it is what the law considers to be either a "covered entity" – namely: a health care provider, health plan, or health care clearinghouse – or a relevant business associate (BA). [79]

This also implies that content such as Facebook or Twitter data, despite their actual use as health indicators, are currently not protected under the HIPAA. Yet, although arguably unlikely, this may change in the future. In addition, scientists should not conflate legal with ethical requirements.

With regard to biomedical research, it has been pointed out that the ethical implications of big data research are, at least partly, uncharted territory. Additional ethical considerations for pharmacoepidemiologic research involving big data are thus needed. This applies to the autonomy of data subjects but also to new corporate stakeholders and public–private partnerships. The latter may not merely involve pharmaceutical companies or device manufacturers. Internet and technology corporations may also play a role and require ethical as well as legal oversight, since they control access to digital data that could further complement pharmacoepidemiology in the future.

Acknowledgment

The authors would like to thank Philip Phan for assistance with literature searches.

References

1 Greene JA. The afterlife of the prescription: sciences of prescription surveillance. In: Greene JA, Watkins ES, eds. *Prescribed: Writing, Filling, Using, and Abusing Prescriptions in Modern America*. Baltimore: Johns Hopkins University Press, 2012.

2 Daemmrich A, Greene JA. From visible harm to relative risk: overcoming fragmented

pharmacovigilance. In: Elhage E, ed. *The Fragmentation of U.S. Health Care: Causes and Solutions*. Oxford: Oxford University Press, 2010, pp. 301–23.
3 Lenz W. Thalidomide and congenital abnormalities. *Lancet* 1962; **279**(7219): 45.
4 Stephens T, Brynne R. *Dark Remedy: The Impact of Thalidomide and Its Revival as a Vital Medicine*. New York: Basic Books, 2001.
5 Kunz W, Keller H, Mückter H. N-Phthalyl-glutaminsäure-imid: experimentelle untersuchungen an einem neuen synthetischen produkt mit sedativem eigenschaften. *Arzneimittelforschung* 1956; **6**: 426–30.
6 Jung H. Klinische erfahrungen mit einem neuen sedativum. *Arzneimittelforschung* 1956; **6**: 430–34.
7 Taussig H. A study of the German outbreak of phocomelia. *JAMA* 1962; **180**: 1106–14.
8 Daemmrich A. A tale of two experts: thalidomide and political engagement in the United States and West Germany. *Social Hist Med* 2002; **15**: 137–58.
9 Carpenter D. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton: Princeton University Press, 2010.
10 Jackson C. *Food and Drug Legislation in the New Deal*. Princeton: Princeton University Press, 1970.
11 Marks H. *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*. Cambridge: Cambridge University Press, 2000.
12 American Medical Association-Boston Collaborative Drug Surveillance Program. Pilot Drug Surveillance Study, Progress Note Number 1, John Adriani Papers (JAP), Box 28, Folder 2, NLM: 1.
13 Kefauver–Harris Amendments, Public Law 87-781, 10 October 1962, 76 Stat. 783.
14 FDA Seminar on Adverse Reactions. 6–9 September 1966, JAP, Box 27, NLM.
15 Federal Register 32, 6 June 1967: 8080–9.
16 Federal Register 33, 11 May 1968: 7077–8.
17 Food and Drug Administration. *Adverse Drug Reaction Reporting*. Montgomery: Food and Drug Administration 1969.
18 Johnson J. Leighton E. Cluff (1923–2004). *Trans Am Clin Climatol Assoc* 2005; **116**: xlv–l.
19 Cluff L, Thornton G, Seidl G. Studies on the epidemiology of adverse drug reactions. I. Methods of surveillance. *JAMA* 1964; **188**: 977.
20 Finney D. The design and logic of a monitor of drug use. *J Chron Dis* 1965: **18**: 77–98.
21 Slone D, Jick H, Borda I, *et al*. Drug surveillance utilizing nurse monitors. An epidemiological approach. *Lancet* 1966; **2**(7469): 901–3.
22 Slone D, Gaetano L, Lipworth L, Shapiro S, Lewis GP, Jick H. Computer analysis of epidemiologic data on effect of drugs on hospital patients. *Public Health Rep* 1969;**84**(1): 40.
23 Jick H, Miettinen OS, Shapiro S, Lewis GP, Siskind V, Slone D. Comprehensive drug surveillance. *JAMA* 1970; **213**(9): 1455–60.
24 Borda I, Slone D, Jick H. Assessment of adverse reactions within a drug surveillance program. *JAMA* 1968; **205**(9): 645–7.
25 Borda I, Jick H, Slone D, Dinan B, Gilman B, Chalmers TC. Studies of drug usage in five Boston hospitals. *JAMA* 1967; **202**(6): 170–4.
26 Shapiro S, Slone D, Lewis GP, Jick H. The epidemiology of digoxin: a study in three Boston hospitals. *J Chron Dis* 1969; **22**(5): 361–71.
27 Jick H, Slone D, Borda IT, Shapiro S. Efficacy and toxicity of heparin in relation to age and sex. *N Engl J Med* 1968; **279**(6): 284.
28 Jick H. Drugs – remarkably nontoxic. *N Engl J Med* 1974; **291**(16): 824–8.
29 Herbst AL, Ulfelder H, Poskanzer D. Adenocarcinoma of the vagina: association of maternal Stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971; **284**(15): 878–81.
30 Jick H, Walker A, Spriet-Pourra C. Postmarketing follow-up. *JAMA* 1979; **242**(21): 2311.

31 *Final Report*, Minutes of Meeting, November 30, 1976.

32 Jick H. The Commission on Professional and Hospital Activities – professional activity study: a national resource for the study of rare illnesses. *Am J Epidemiol* 1979;**109**(6): 625–7.

33 Bjornson DC, Serradell J, Hartzema AG. Drug utilization – measurement, classification and methods. In: Hartzema AG, Porta MS, Tilson HH, eds. *Pharmacoepidemiology: An Introduction*. Cincinnati: Harvey Whitney Books, 1998, pp. 131–60.

34 World Health Organization. *Consumption of Drugs: Report on a Symposium* (EURO 3102). Copenhagen: Regional Office for Europe, 1970.

35 Bergman U. The history of the Drug Utilization Research Group in Europe. *Pharmacoepidemiol Drug Saf* 2006; **15**(2): 95–8.

36 Ekedahl A, Libdeck J, Lithman T, Noreen D, Melendar A. Benzodiazepine prescribing patterns in a high-prescribing Scandinavian community. *Eur J Clin Pharmacol* 1933; **44**: 141–6.

37 Bergman U, Sjoqvist F. Measurement of drug utilization in Sweden: methodological and clinical implications. *Acta Med Scand* 1984; **683**: 15–22.

38 Ramirez I. Navigating the maze of requirements for obtaining approval of non-interventional studies (NIS) in the European Union. *Ger Med Sci* 2015; **13**: Doc 21.

39 Regulation (EU) No 1235/2010 (2010, December 15). *Official Journal of the European Union*, 348: 1–16. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:348:0001:0016:EN:PDF (accessed May 17, 2019).

40 Mullard, A. Mediator scandal rocks French medical community. *Lancet* 2011; **377**(9769): 890–2.

41 Kimura T, Matsushita Y, Yang YH, Choi NK, Park BJ. Pharmacovigilance systems and databases in Korea, Japan, and Taiwan. *Pharmacoepidemiol Drug Saf* 2011; **20**: 1237–45.

42 Lai EC, Man KK, Chaiyakunapruk N, *et al*. Databases in the Asia-Pacific Region: the potential for a distributed network approach. *Epidemiology* 2015; **26**(6): 815–20.

43 Huang YL, Moon J, Segal JB. A comparison of active adverse event surveillance systems worldwide. *Drug Saf* 2014; **37**(8): 581–96.

44 AsPEN Collaborators, Andersen M, Bergman U, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; **22**(7): 700–4.

45 Pratt B, Van C, Cong Y, *et al*. Perspectives from South and East Asia on clinical and research ethics. *J Empir Res Hum Res Ethics* 2014; **9**(2): 52–67.

46 Merz JF. Introduction: a survey of international ethics practices in pharmacoepidemiology and drug safety. *Pharmacoepidemiol Drug Saf* 2001; **10**(7): 579–81.

47 Urushihara H, Parmenter L, Tashiro S, Matsui K, Dreyer N. Bridge the gap: the need for harmonized regulatory and ethical standards for postmarketing observational studies. *Pharmacoepidemiol Drug Saf* 2017; **26**(11): 1299–306.

48 Bothwell L. The emergence of the randomized controlled trial: origins to 1980. Doctoral thesis, Columbia University, 2014.

49 Parkin L, Paul C. Public good, personal privacy: a citizens' deliberation about using medical information for pharmacoepidemiological research. *J Epidemiol Commun Health* 2011; **65**(2): 150–6.

50 Fairchild A, Bayer R, Colgrove J. *Searching Eyes: Privacy, The State, and Disease Surveillance in America*. Berkeley: University of California Press, 2007.

51 Leufkens HG. Privacy issues in pharmacoepidemiology: the importance of weighing costs and benefits. *Pharmacoepidemiol Drug Saf* 2001;**10**(7): 659–62.

52 Office for Human Research Protections, United States Department of Health and

Human Services. 45 Code of Federal Regulations Part 46, 2009. www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html.

53 Khoury A, Storm BL, Kimmel SE, Hennessy S. Bioethical issues in pharmacoepidemiologic research. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Oxford: Wiley-Blackwell, 2012, pp. 633–5.

54 European Medical Agency (n.d.). Good pharmacovigilance practices. Retrieved from www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000345.jsp (accessed May 17, 2019).

55 Verhamme K, Sturkenboom M. Study designs in pediatric pharmacoepidemiology. *Eur J Clin Pharmacol* 2011; **67**(Suppl 1): 67–75.

56 Zeitoun JD, Ross JS, Ignacio A, *et al*. Postmarketing studies for novel drugs approved by both the FDA and EMA between 2005 and 2010: a cross-sectional study. *BMJ Open* 2017; **7**(12): e018587.

57 Hopf YM, Bond CB, Francis JJ, Haughney J, Helms PJ. Linked health data for pharmacovigilance in children: perceived legal and ethical issues for stakeholders and data guardians. *BMJ Open* 2014; **4**(2): e003875.

58 Cristea IA, Gentili C, Pietrini P, Cujpers P. Sponsorship bias in the comparative efficacy of psychotherapy and pharmacotherapy for adult depression: meta-analysis. *Br J Psychiatry* 2017; **210**(1): 16 LP–23.

59 Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017; 2: MR000033.

60 International Society of Pharmacoepidemiology Public Policy Committee. Guidelines for Good Pharmacoepidemiology Practice (GPP). *Pharmacoepidemiol Drug Saf* 2016; **25**: 188–91.

61 ENCePP. (n.d.). The European Union electronic Register of Post-Authorisation Studies (EU PAS Register). www.encepp.eu/encepp_studies/indexRegister.shtml (accessed May 17, 2019).

62 Bothwell L, Greene J, Podolsky S, Jones D. Assessing the gold standard – lessons from the history of RCTs. *N Engl J Med* 2016; **374**: 2175–81.

63 Regulation (EU) No 536/2014 (2014, May 27). *Official Journal of the European Union*, **158**: 1–76. https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf (accessed May 17, 2019).

64 Cunningham K. Clinical Trials Regulation (EC) No. 536/2014 Update on the EU Portal and Database. www.ema.europa.eu/docs/en_GB/document_library/Presentation/2017/05/WC500228179.pdf (accessed May 17, 2019).

65 Abou-El-Enein M, Schneider C. Deciphering the EU clinical trials regulation. *Nature Biotechnol* 2016; **34**: 231–3.

66 Dittrich C, Negrouk A, Casali P. An ESMO-EORTC position paper on the EU clinical trials regulation and EMA's transparency policy: making European research more competitive again. *Ann Oncol* 2015; **26**(5): 829–32.

67 Westra A, Bos W, Cohen A. New EU clinical trials regulation: needs a few tweaks before implementation. *BMJ*, 2014; **348**: g3710.

68 ENCePP. Considerations on the Definition of Non-Interventional Trials under the Current Legislative Framework ("Clinical Trials Directive" 2001/20/EC). www.encepp.eu/publications/documents/ENCePPinterpretationofnoninterventionalstudies.pdf (accessed May 17, 2019).

69 ENCePP. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. www.encepp.eu/standards_and_guidances/methodologicalGuide9.shtml (accessed May 17, 2019).

70 Ephross S, Mitchell A, Sacks S, Straus W, Stummer T. Principles for effective academia-industry collaboration in pharmacoepidemiology. www.pharmacoepi.org/pub/?id=74d2dc06-de25-47ab-2dce-063705086d1f (accessed May 17, 2019).

71 Drug Safety Research Unit. www.dsru.org (accessed May 17, 2019).

72 Matthews G, Williams N. Post Brexit crystal ball gazing: what the future holds for phase I clinical trials in the UK. *Clinical Risk* 2017; **22**(5–6): 97–104.
73 Walker R. Leaving the EU: Implications and opportunities for science and research (letter). www.parliament.uk/documents/commons-committees/science-technology/170921-Robin-Walker-to-Norman-Lamb-DExEU%20letter.pdf (accessed May 17, 2019).
74 Sethi N. The promotion of data sharing in pharmacoepidemiology. *Eur J Health Law* 2014; **21**(3): 271–96.
75 Wettermark B. The intriguing future of pharmacoepidemiology. *Eur J Clin Pharmacol* 2013; **69**(Suppl 1): 43–51.
76 Freifeld CC, Brownstein JS, Menone CM, *et al*. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf* 2014; **37**(5): 343–50.
77 Science and Technology Committee. The big data dilemma. www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/46809.htm#_idTextAnchor035 (accessed May 17, 2019).
78 Jardine J, Fisher J, Carrick B. Apple's ResearchKit: smart data collection for the smartphone era? *J Roy Soc Med* 2015; **108**(8): 294–6.
79 Electronic Frontier Foundation. Medical Privacy. www.eff.org/issues/medical-privacy (accessed May 17, 2019).

Further Reading

Abou-El-Enein M, Schneider C. Deciphering the EU clinical trials regulation. *Nature Biotechnol* 2016; **34**: 231–3.
Hedgecoe A. Scandals, ethics, and regulatory change in biomedical research. *Sci Technol Human Values* 2016; **42**(4): 577–99.
International Society of Pharmacoepidemiology Public Policy Committee. Guidelines for Good Pharmacoepidemiology Practice (GPP). *Pharmacoepidemiol Drug Saf* 2016; **25**: 188–91.
Pratt B, Van C, Cong Y, *et al*. Perspectives from South and East Asia on clinical and research ethics. *J Empir Res Hum Res Ethics* 2014; **9**(2): 52–67.
Ramirez I. Navigating the maze of requirements for obtaining approval of non-interventional studies (NIS) in the European Union. *German Med Sci* 2015; **13**. www.ncbi.nlm.nih.gov/pmc/articles/pmid/26633964 (accessed April 22, 2019).
Urushihara H, Parmenter L, Tashiro S, *et al*. Bridge the gap: the need for harmonized regulatory and ethical standards for postmarketing observational studies. *Pharmacoepidemiol Drug Saf* 2017; **26**: 1299–306.
Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. *PLoS Comput Biol* 2015; **11**(2). http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003904 (accessed April 22, 2019).
Wettermark B. The intriguing future of pharmacoepidemiology. *Eur J Clin Pharmacol* 2013; **69**(1): 43–51.

32

The Use of Randomized Controlled Trials in Pharmacoepidemiology

Robert F. Reynolds[1], Samuel M. Lesko[2], Nicolle M. Gatto[1], Tjeerd P. van Staa[3], and Allen A. Mitchell[4]

[1] *Department of Epidemiology, Global Medical Value, Evidence and Outcomes, GlaxoSmithKline, Upper Providence, PA, USA*
[2] *Northeast Regional Cancer Institute, Scranton, PA, USA*
[3] *Health eResearch Centre, University of Manchester, Manchester, UK*
[4] *Slone Epidemiology Center at Boston University, Boston, MA, USA*

When properly conducted, randomized controlled trials (RCTs) are considered the gold standard for demonstrating the efficacy and safety of a medicine for regulatory approval. During the premarketing phases of drug development, RCTs typically involve highly selected subjects and in the aggregate include at most a few thousand patients. These studies are designed to be sufficiently large to provide evidence of a beneficial clinical effect and to exclude large increases in risk of common adverse clinical events. Premarketing trials are rarely large enough to detect small differences in the risk of common adverse events or to estimate reliably the risk of rare events, whether serious or trivial (see Chapters 1 and 4). Quantification of these potentially important rare risks requires large studies, which typically are conducted after a drug is marketed. Because of design complexity and costs, large controlled trials are not generally considered for the postmarketing safety evaluation of drugs.

Rather, observational designs are commonly used to evaluate the safety of medicines post approval. They may be the only means to study large populations under routine clinical conditions or to evaluate a medicine's association with rare events or long-latency outcomes. Observational studies are therefore a logical complement to preapproval safety data collected from RCTs, expanding the knowledge about a medicine's safety profile as it is used in larger and more diverse patient populations. It is now generally accepted that data on a drug's use among real-world populations strengthen its evaluation, whether through organized research efforts or voluntary reporting systems [1–4], and pharmacoepidemiologic studies as a condition of regulatory approval for new medicines have increased over the past 15 years [5,6].

Despite notable advances in observational methods for assessing medicine safety, findings from observational studies are frequently contested as the basis for regulatory and clinical decisions. One important reason is that epidemiologic studies of medication exposures and their effects have difficulty measuring and controlling for confounding in general, and confounding by indication for drug use (and/or severity of disease) specifically (see Chapter 33)

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

[7–9]. Concerns about uncontrolled confounding are central to debates about the interpretation of observational study findings and have complicated the assessment of many potential medicine–adverse outcome associations [9–21].

Nonetheless, the options for investigating real-world safety, or the safety of a medicine as it is actually prescribed by physicians and used by patients once on the market, are limited [21]. Postapproval RCTs are ethically and logistically difficult to conduct, are not appropriate to address many safety issues (e.g., rare or longer-term outcomes), and are unlikely to provide information on the safety of medicines in real-world populations. Thus, studies that are observational and follow patients with minimal interference are often the only means by which to answer clinically meaningful questions about a marketed medicine's safety and to study patients in settings that generalize to real-world medicine use.

The ideal design for a postapproval safety study is one that minimizes the potential for bias yet is still relevant to real-world clinical practice. A design that in principle merges the ideal characteristics of the randomized controlled trial (randomization) with those of an observational epidemiology study (follow-up with minimal intervention) is the large simple trial (LST) [22–24]. LSTs are characterized by large sample sizes, often in the thousands; broad entry criteria consistent with the approved medication label; randomization based on equipoise, that is, neither physician or patient believes that one treatment option is superior; minimal data requirements such as a questionnaire or case report form usually collecting data in only a few pages with questions limited to key variables that are typically collected at routine clinical care visits; objectively measured endpoints (e.g., death, hospitalization); follow-up that minimizes interventions or interference with normal clinical practice; follow-up of all patients regardless of whether they discontinue randomized medication; and intent-to-treat (ITT) analysis examining the entire population of randomized subjects according to the treatment group to which they were initially randomized.

Although the LST design shares a key design component, randomization, with an RCT, it is distinguished by its intent to minimize interference with usual medical care. In the typical RCT, the intent is restriction and control to create experimental conditions at baseline and over the course of the study. In contrast, in the LST the intent is to create balance of baseline characteristics but to then follow patient outcomes using observational methods. Practically, this means that endpoint definition, physician and patient recruitment, drug delivery, data collection, allowance for treatment discontinuation, concomitant drug use, and patient and site monitoring in a LST are operationally different from that of a RCT. Recent studies referred to as pragmatic clinical trials (PCTs) typically fall somewhere in between these approaches, where the goal is to introduce one or more pragmatic elements into the design but with substantial protocol-required follow-up and testing outside usual care practice. We view the LST as a type of PCT, at the far end of a continuum of pragmatic randomized designs, where the goal of the LST is to adopt as many pragmatic elements as possible to mimic usual care practice while protecting study validity.

Most calls for the use of the LST design have, until recently, primarily focused on the need for studies of clinically important therapeutic or preventive effects of interventions to inform clinical and health policy decisions [23,24]. LSTs have been employed to study a range of real-world benefits, from interventions for the treatment and prevention of cardiovascular outcomes [25–29] to a comparison of antiretroviral treatment strategies for HIV-positive patients [30]. In the 1990s, two research groups (one of which included SL and AM) described its use for safety, each using a single LST they had completed as a case study to infer general lessons for future studies [31,32]. More recently,

the US Food and Drug Administration (FDA) [33] and the European Medicines Agency (EMA) [34] acknowledged the potential advantages for postmarketing safety evaluation due to its unique design characteristics.

Clinical Problems to be Addressed by Pharmacoepidemiologic Research

Pharmacoepidemiologic methods are classically used to quantify risks and benefits of medicines that could not be adequately evaluated in studies performed during the premarketing phase of drug testing. In this chapter, we focus on the role of LSTs for assessing the risks of medications; the same principles, however, are applicable to the postmarketing study of benefits.

As noted in the introduction, if there are questions about the safety of a drug after it has been licensed, large observational studies are typically used to satisfy the sample sizes needed to identify (or rule out) the relevant risks. The respective strengths and weaknesses of these designs are discussed elsewhere in this volume (see Chapter 3). Potential confounding is a major concern for virtually every observational study, and uncontrolled or incompletely controlled confounding can easily account for modest associations between a drug and an adverse clinical event. For example, in the relation between phenylpropanolamine and cerebrovascular disease, obesity increases both the likelihood of exposure to the drug and the risk of a cerebrovascular accident; thus, body weight must be controlled in any analysis of this association. The challenge to the pharmacoepidemiologist is to recognize those factors that represent potential confounders, validly measure them and then control for their effects. To do so requires that the relevant information be included in the data to be analyzed, but information on important confounding factors is frequently incomplete or unavailable. Surrogate variables are often used (e.g., years of education to reflect socioeconomic status) but these may be poor measures of the underlying confounding factor and their control therefore may not eliminate confounding.

In observational studies, weak associations deserve particular attention. Although there are important exceptions, the general view is that the stronger the association, the more likely it is that the observed relationship is causal. This is not to say that a weak association (e.g., a relative risk ≤1.5) can never be causal; rather, it is more difficult to be certain of it because such associations, even if statistically significant, can easily be an artifact of confounding. As an example, consider an analysis where socioeconomic status is a potential confounder and education is used as a surrogate for this factor. Because the relation between years of education completed (the surrogate) and socioeconomic status (the potential confounder) is, at best, imperfect, analyses controlling for years of education can only partially control for confounding. This leads to the familiar caveat in reports of observational studies, "…residual confounding may account for the observed association."

In pharmacoepidemiologic research, the phenomenon known as confounding by indication (also referred to as indication bias, channeling, confounding by severity, or contraindication bias) is a common methodologic problem (see also Chapters 33 and 43). According to Slone *et al.*, confounding by indication exists when "patients who receive different treatments … differ in their risk of adverse outcomes, independent of the treatment received" [35]. In general, confounding by indication occurs when an observed association between a drug and an outcome is due to the underlying illness (or its severity) and not to any effect of the drug. Put another way, confounding by indication occurs when the risk of an adverse event is related to the *indication* for medication use but not the

use of medication itself. As with any other form of confounding, one can, in theory, control for its effects if one can reliably measure the severity of the underlying illness. In practice, this is not easily done (see Chapters 33 and 43).

When confronted with the task of assessing the safety of a marketed drug product, the pharmacoepidemiologist must evaluate the specific hypothesis to be tested, estimate the magnitude of the hypothesized association, and determine whether confounding by indication is possible. If incomplete control of confounding is likely, it is important to recognize the limitations of observational research designs and consider conducting a randomized study design such as the LST. There is nothing inherent about a randomized design that precludes a pharmacoepidemiologist from designing and carrying it out. On the contrary, the special skills of a pharmacoepidemiologist can be very useful in performing large-scale randomized trials that use epidemiologic follow-up methods.

Methodologic Problems to be Solved by Pharmacoepidemiologic Research

Large simple trials may be the best solution when it is not possible to completely control confounding by means other than randomization. LSTs are really just very large randomized trials made simple by reducing data collection to the minimum needed to test only a single hypothesis (or at most a few hypotheses). Randomization of treatment assignment is the key feature of the design, which controls for confounding by known and unknown factors. The large study size provides the power needed to evaluate small risks, either absolute or relative. Table 32.1 highlights the design aspects that differentiate LSTs from RCTs.

How Simple is Simple?

Yusuf *et al*. have suggested that very large randomized studies of treatment-related mortality need collect only data concerning the vital status of participants at the conclusion of the study [22]. Because the question of drug safety frequently concerns outcomes less severe than mortality, these ultra-simple trials may not be sufficient. Hasford has suggested a somewhat less restrictive approach to data collection, in which "large trials with lean protocols" include only *relevant* baseline, follow-up, and outcome data [31]. Collecting far less data than is common in the usual RCT is the key feature of both approaches. With simplified protocols that take advantage of epidemiologic follow-up methods, very large trials can be conducted to test hypotheses that are relevant to clinical practice.

Concealed (Masking) Versus Known Treatment Assignment

Randomized controlled trials conceal treatment assignment to minimize detection bias and, in most cases, are only considered to be unbiased if all parties (patients, healthcare providers, and investigators) are unaware of the treatment assigned. Masking is particularly important when the outcome is subjective and/or investigators have views about the medications under study that may lead to differential testing, detection, and reporting of outcomes. Even in RCTs, however, concealing medications is not a panacea for unbiased outcome assessment. Masking has methodological limitations, ranging from investigators or patients guessing treatment allocation due to differential improvement in symptoms or outcomes to failures in the taste, texture or color of the random treatment assignments. In addition to the bias this may introduce for outcome assessment, it may also create selection bias if investigators are able to predict the likelihood of an upcoming treatment allocation. For this reason, RCT investigators use

Table 32.1 Typical design characteristics of a large simple trial (LST) compared to those of a randomized controlled trial (RCT).

Design characteristic	LST	RCT
Randomization	Yes	Yes
Medicine assignment	Concealed if feasible, assignment may be known, dose adjustment permitted	Concealed
Sample size	Larger (thousands)	Smaller (hundreds)
Inclusion criteria	Broad (e.g., per approved medicine label)	Narrow (e.g., excludes patients with co-morbid conditions, using multiple medications, pregnant women, elderly)
Questionnaire/case report form (CRF)/interview	Minimal, brief	Complex, lengthy
Endpoints	Hard endpoints (mortality, hospitalization or life-threatening events)	Virtually any
Required patient visits and procedures	Few, if any; follows normal practice schedule and assessments	Yes, frequent; visits and tests far greater than expected in clinical practice
Primary source of investigators or enrolling physicians	Primary care provider/ community based	Clinical research/academic centers
Site monitoring	Minimal	Frequent
Followed after randomized treatment discontinued	Yes	No, or for limited duration post discontinuation (e.g., 30 days)
Primary analytic method	Intention to treat (ITT)	ITT

methods for assessing the success of concealment and real-time central randomization to address the potential for selection bias.

The value of concealing medications to improve the validity of randomized study designs has been established in countless RCTs. Masking treatment allocation is possible and has been proven to be successful in LSTs (as described later). Therefore, pharmacoepidemiologists should start by evaluating whether concealment is feasible for their trial design. Given practical considerations, many LSTs will not be able to implement masking due to the costs and infrastructure complexity or other factors such as study duration. Regulators and policy makers may also find concealment undesirable, particularly when balancing the need for evidence that reflects real-world clinical practice. In fact, for some LSTs, the primary goal is to only compare treatment strategies in usual care. This requires minimal intervention so that physicians and patients behave as they normally would when taking a particular medicine. In these LSTs, investigators are less concerned with the causal effects of a medicine (as would be measured in a controlled clinical trial) and more concerned with the safety outcomes that occur due to the awareness of treatment allocation and the associated, and potentially differential, practices employed by physicians and patients as a result. LSTs that do not use concealment must address the potential

for bias and will be strengthened by comparing hard clinical outcomes or requiring similar medical work-up/testing across treatment arms for study outcomes.

Ultimately, whether to conceal medicine assignment is a critical design decision for the pharmacoepidemiologist that depends on numerous factors, including the research question, outcomes under study, patient population, preferences of patients, and the preferred balance between validity and reflecting real-world clinical practice.

Power/Sample Size

Study power is related to the number of events observed during the course of the study, which in turn is a function of the incidence rate for the event, sample size, persistence to the study treatment, and duration of observation (or follow-up). Power requirements can be satisfied by studying a population at high risk, enrolling a large sample size, and/or conducting follow-up for a prolonged period. The appropriate approach will be determined by considering the goal of the study and the hypothesis to be tested. Allergic or idiosyncratic events may require a very large study population, and events with latency periods must include duration of follow-up consistent with the hypothesized timing of symptom onset. The decision to study high-risk populations must be balanced with the need for generalizability. For example, while an elderly population may meet criteria for high risk of cardiovascular events, a study limited to this group would be inappropriate to assess the risk of these events in younger adults or children.

Data Elements

The data collection process can be kept simple by restricting the study to a few primary endpoints that satisfy the study hypothesis, are objective, are easily identified, and are verifiable. Some researchers may need to overcome their predisposition to comprehensive data collection when it comes to secondary outcomes (i.e., those that do not directly relate to the study hypothesis), as these must be ignored to eliminate unnecessary effort and complexity. Of critical importance, because confounding is controlled by randomization, data on all potential confounders need not be collected. Rather, a few basic demographic variables can be collected at enrollment in order to characterize the population studied and allow the investigators to confirm that effective randomization was achieved.

Data Collection

The data collection process itself can be streamlined. Follow-up data can be collected directly from participants via, for example, mailed questionnaires, telephone interviews, online using a secure website, or through mobile devices. Because the study will be limited to clear and objective outcomes which can be confirmed by medical record review or other means, reports from study participants, family care providers or, for some patient populations, healthcare providers can be an appropriate source of follow-up data. Other sources of follow-up data could include electronic medical records (e.g., for studies among subscribers of a large health care organization where it is likely that important outcomes will be recorded) or vital status records for fatal outcomes (e.g., the US National Death Index).

The primary advantage of this simplicity is that it allows very large groups of study participants to be followed at reasonable cost. The trade-off is that a simple trial cannot answer all possible questions about the safety of a drug but must be limited to testing, at most, a few related hypotheses.

When is a LST Appropriate?

Large simple trials are appropriate when all the conditions in Box 32.1 apply.

Box 32.1 Conditions appropriate for the conduct of a large simple trial

1) The research question is important
2) Genuine uncertainty exists about the likely results
3) Confounding by indication is likely
4a) The absolute risk is small
or
4b) The relative risk is small, regardless of the absolute risk

Important Research Question

Although a simple trial will likely cost considerably less per subject than a typical premarketing clinical trial, the total cost of a large study (in money and human resources) will still be substantial. The cost will usually be justified only when there is a clear need for a reliable answer to a question concerning the risk of a serious outcome.

Uncertainty Must Exist

An additional condition has been referred to as the "uncertainty principle." This was originally described by Gray *et al*. as a simple criterion to assess subject eligibility in LSTs [36]. It states that "… both patient and doctor should be *substantially uncertain* about the appropriateness, for this particular patient, of each of the trial treatments. If the patient and doctor are *reasonably certain* that one or other treatment is inappropriate then it would not be ethical for the patient's treatment to be chosen at random" (italic in the original). Very large randomized safety trials are justified only when there is true uncertainty about the risk of the treatment in the population. Apart from considerations of benefit, it would not be ethical to subject large numbers of patients to a treatment that was reasonably believed to place them at increased risk, however small, of a potentially serious or permanent adverse clinical event. The concept of uncertainty can thus be extended to include a global assessment of the combined risks and benefits of the treatments being compared. One treatment may be known to provide therapeutic benefits that are superior to an alternative, but it may be unknown whether the risks of important side effects outweigh the therapeutic advantage.

Power and Confounding

Large simple trials will only be needed if (a) the *absolute* risk of the study outcome is small and there are concerns about confounding by indication, or (b) the *relative* risk is small (in which case, there are inherent concerns about residual confounding) [37]. By contrast, LSTs would not be necessary if the *absolute* risk were large, because premarket or other conventional RCTs should be adequate, or where confounding by indication is not an issue, because observational studies would suffice. Also, if the *relative* risk were large (and confounding by indication is not a concern), observational study designs are likely appropriate.

When is an LST Feasible?

Large simple trials are feasible when all the conditions in Box 32.2 are met.

Box 32.2 Conditions which make a large simple trial most feasible

1) The study question can be expressed as a simple testable hypothesis
2) The treatment to be tested is simple (uncomplicated)
3) The outcome is objectively defined (e.g., hospitalization, death)
4) Epidemiologic follow-up methods are appropriate
5) A cooperative and motivated population is available for study

Simple Hypothesis

Large simple trials are best suited to answer focused and relatively uncomplicated questions. For example, an LST can be designed to test the hypothesis that the risk of hospitalization for any reason, or for acute gastrointestinal bleeding, is increased in children treated with ibuprofen. However, it may not be possible for a single LST to answer the much more general question, "Is ibuprofen safe with respect to all possible outcomes in children, whether or not they require medical attention?" Similarly, in a study of the clinical relevance of QTc prolongation, a LST can test the hypothesis that the risk of mortality is increased from this effect, but will not be able to evaluate the increased risk of ventricular arrhythmias or its specific forms, such as torsade de pointes.

Simple Treatments

Simple therapies (e.g., a single drug at a fixed dose for a short duration) are most amenable to study with LSTs. They are likely to be commonly used, so that it will be easy to enroll large numbers of patients, and the results will be applicable to a large segment of the population. Complex therapeutic protocols are difficult to manage, reduce patient adherence, and by their very nature may not be compatible with the simple trial design.

Objective and Easily Measured Outcomes

The outcomes to be studied should be objective, easy to define ("simple"), and easy to recall. An example might include hospitalization for acute gastrointestinal bleeding. Study participants may not recall the details of a hospital admission, but they likely will recall the fact that they were admitted, the name of the hospital, and at least the approximate date of admission. Medical records can be obtained to document the details of the clinical events that occurred. Events of this type can be reliably recorded using epidemiologic follow-up methods (e.g., questionnaires, telephone interviews, online surveys, mobile device applications, hospital discharge diagnosis codes, or linkage with public vital status records). On the other hand, clinical outcomes which can be reliably detected only by detailed in-person interviews, physical examinations, or extensive physiologic testing are not as amenable for study in simple trials. For example, an LST concerned about the QTc prolonging potential of a medicine will focus on the clinically relevant and measurable sequelae rather than on performing routine ECGs for study participants.

Cooperative Population

Particularly in LSTs, a cooperative and motivated study population greatly increases the probability of success. Striking examples are the large populations in the Physicians' and Women's Health Studies; the success of these studies is at least partly due to the willingness of large numbers of knowledgeable health professionals to participate [38,39]. Because of the participants' knowledge of medical conditions and symptoms and participation in the US healthcare system, relatively sophisticated information could be obtained using mailed questionnaires, and even biologic samples could be collected. In another example, success of the Boston University Fever Study (described later) was also largely due to parents whose motivation and cooperation were encouraged by their private physicians who had invited them to participate in the study [40]. Similarly, in the ZODIAC LST (also described later) among patients with schizophrenia, the motivation of psychiatrists and family caregivers around a clinically important question led to robust one-year follow-up despite a patient population difficult to follow, even short term (e.g., in 12–16-week RCTs) [41,42].

Logistics of Conducting a LST

An LST may be appropriate and feasible, but it will only succeed if all logistical aspects of the study are kept simple as well. In general, LSTs will involve an oversight body, sometimes organized as a scientific steering committee composed of epidemiologic, statistical, and clinical experts who are responsible for the scientific conduct of the study, as well as a central data coordinating facility, and a network of enrollment sites (e.g., offices of collaborating physicians or other healthcare providers). Healthcare professionals (e.g., physicians, nurse practitioners, and pharmacists in private practice or members of large healthcare organizations) can participate by recruiting eligible patients. Alternative methods to identify and enroll eligible subjects (e.g., direct mailings to professional groups, print or online ads, emails and mobile phone text messages) may be appropriate for some studies.

Because success depends on the cooperation of multiple healthcare providers and a large number of patients, it is best to limit the demands placed on each practitioner (or his/her clinical practice). One approach is to have the practitioner identify eligible subjects, obtain permission to pass their names to a central study staff, and leave to the study staff the task of explaining study details, enrollment, and obtaining informed consent. Another approach is to provide comprehensive training prior to site initiation followed by support to local administrative staff throughout the course of the study, particularly for research-naive and inexperienced sites. Obtaining informed consent, baseline data, and the medicine assignment is best handled during the course of a single visit.

To facilitate patient recruitment and to maximize generalizability of the results, minimal restrictions should be placed on patient eligibility. As Gray *et al*. have said, "Any obstacle to simplicity is an obstacle to large size, and the wider the range of the patients studied, the wider the generalizability of the results will be" [36]. Patients with a medical contraindication or known sensitivity to either the study or control drug should not, of course, be enrolled, but other restrictions should be kept to a minimum. Ideally, the restrictions are only those that apply in a usual care clinical setting, that is, those described in the approved medicine label.

Simple informed consent and registration documents should be completed. Registration of study subjects can also be accomplished online using a secure internet connection (or potentially through secure mobile phone applications) to the coordinating center, which allows for immediate confirmation of eligibility and randomization. Substantial bias can be introduced if either physician or patient can choose not to participate after learning (or guessing) which treatment the patient has been assigned.

Particularly in studies requiring a long duration of medication use, validity may be seriously compromised by poor adherence with the treatment regimen. A run-in period prior to randomization can be used to identify patients who are unable or unwilling to adhere to a chronic treatment regimen and are likely to drop out of the study. During the run-in period, eligible subjects are given a "test" medication and their compliance with the protocol is assessed. Patients who cannot comply with the protocol are withdrawn. Patients who remain in the study are likely to be highly adherent, so that relatively few will drop out after randomization. Depending on the characteristics of drugs under study, either the active drug or the control may be preferable for the run-in period. In the Physicians' Health Study, for example, the study drug aspirin was used for the run-in period to identify subjects who could not tolerate the gastrointestinal side effects of the drug [38]. A run-in period may lead to a study population that no longer reflects the real-world users of a medicine. This approach should therefore be used sparingly and only when the conditions justify it.

Importance of Complete Follow-Up

Because dropouts and losses to follow-up may not be random but rather may be related to adverse treatment effects, it is important to make every effort to obtain follow-up data on all enrolled subjects. For example, a study that has follow-up data on even tens of thousands of patients may not be able to provide a valid answer to the primary study question if this number represents only half of those randomized. The duration of the follow-up period can affect the completeness of follow-up data collection. If the duration of follow-up is too short, important outcomes may be missed (i.e., they may not be diagnosed until after the end of the follow-up period). On the other hand, as the length of the follow-up period increases, the number lost to follow-up or exposed to the alternate treatment (contaminated exposure) increases. In the extreme, a randomized trial becomes a cohort study because of selective dropouts in either or both of the treatment arms.

Beyond choosing a motivated and interested study population, investigators can minimize losses to follow-up by maintaining regular contact with study participants. Regular mailings of medication supplies or prescriptions filled at a local pharmacy that are reimbursed by the study, a study newsletter, or email and mobile phone text reminders can be helpful, and memory aids such as medication calendar packs or other devices may help maintain compliance with chronic treatment schedules. In addition, follow-up data collection itself can help maintain contact with study participants.

Follow-Up Data Collection

An important element of a successful LST is that the burden to healthcare providers for follow-up data collection is minimized. Healthcare providers cannot be expected to consistently obtain substantial amounts of follow-up data from large numbers of subjects. However, the subject's clinician may, with subject permission, provide limited follow-up data (e.g., vital status, occurrence of a primary outcome) or current contact information for the occasional patient who would otherwise be lost to follow-up. A mailed or electronically administered questionnaire, supplemented by telephone follow-up when needed, may work well, although the best means of communication will likely vary by physician specialty and practice size. The response rate will likely be greatest if the questions are simple and direct and the time required to complete the questionnaire is limited. With appropriate permissions, medical records can be reviewed to verify important outcomes, such as rare adverse events; the work needed to obtain and abstract the relevant records should be manageable. In addition, a search of public records (e.g., the National Death Index in the US) can identify study subjects who have died during follow-up.

Analysis

Primary Analysis

Analyses of the primary outcomes are usually straightforward and involve a comparison of incidence rates between the treatment and control groups. Under the assumption that confounding has been controlled by the randomization procedure, complex multivariate analyses are not necessary (and may not be possible because only limited data on potential confounders are available). Descriptive data collected at enrollment should be analyzed by treatment group to verify balance of potential confounders; any material differences between treatment groups suggest a failure of randomization.

Subgroup Analyses

It is important to remember that confounding factors will be distributed evenly only among groups that were randomized; subgroups which are not random samples of the original randomization groups may not have similar

distributions of confounding factors. For example, participants who have remained in the study (i.e., have not dropped out or been lost to follow-up) may not be fully representative of the original randomization groups and may not be comparable with respect to confounders. Despite all efforts, complete follow-up is rarely achieved, and because only the original randomization groups can be assumed to be free of confounding, at least one analysis involving all enrolled study subjects (i.e., an intention-to-treat analysis) should be performed. Also, unless a stratified randomization scheme was used, one cannot be certain that unmeasured confounding variables will be evenly distributed in subgroups of participants, and the smaller the subgroup, the greater the potential for imbalance. Therefore, subgroup analyses can be subject to the same limitations as observational studies (i.e., the potential for uncontrolled confounding).

Data Monitoring/Interim Analyses

Because of the substantial commitment of resources and large number of patients at risk for adverse outcomes, and the interventional nature of the study, it is appropriate for a data monitoring committee (DMC), independent of other scientific and operational study committees and infrastructure, to monitor the accumulating data over the course of the study. The study may be ended prematurely if participants experience unacceptable risks, if the hypothesis can be satisfactorily tested earlier than anticipated, or if it becomes clear that a statistically meaningful result cannot be achieved, even if the study were to be completed as planned.

LSTs in Practice: Two Case Studies

To illustrate these design principles, as well as how they are operationalized, the rationale, design, and results from two completed LSTs are described below.

Boston University Fever Study (BUFS)

Ibuprofen is a nonsteroidal antiinflammatory drug (NSAID) that is widely used among adults in the US, first by prescription and then as an over-the-counter (OTC) drug. In 1989, ibuprofen suspension was licensed as a prescription product for fever control in children, since premarketing studies in children established that it was appropriate for use under a physician's supervision. However, events known to occur in adults using ibuprofen, such as acute gastrointestinal bleeding, acute renal failure, and anaphylaxis, were either not observed at all during the relatively small premarketing trials in children or occurred so infrequently that it was not possible to obtain reliable estimates of the risk. Thus, whether these events, which affect adults, might rarely be caused by ibuprofen in children was unknown. In addition, it was at least theoretically possible that Reye syndrome (a toxic encephalopathy in children associated with another NSAID, aspirin) might be associated with ibuprofen use in children. Other events, possibly unique to children, might also be associated with this drug. Thus, premarketing studies were unable to exclude even a substantially increased risk of rare but important and serious adverse reactions.

Once available OTC, pediatric ibuprofen was likely to be widely used for the treatment of fever, which is typically a minor and self-limited condition. Given the generally benign nature of this indication, it was reasonable to require greater assurance of safety than might be expected for a drug used to treat a life-threatening illness. Further, an effective antipyretic with an excellent record of safety in children, acetaminophen, had been available OTC in the US for more than 20 years. For these reasons, the US FDA required additional data concerning the risk of rare but serious adverse events before it would approve pediatric ibuprofen for OTC sale.

The circumstances surrounding ibuprofen use in 1989–90 raised serious concern that

observational studies could not adequately control confounding. Specifically, prior to the availability of pediatric ibuprofen, febrile children in the US received no antipyretic or were given acetaminophen, which was generally considered safe by both physicians and parents. On the other hand, because ibuprofen was available only by prescription, treatment with this drug required contact with a physician. In addition, for fever less than 102.5 °F, the recommended dose of prescription ibuprofen was 5 mg/kg, whereas for fever of 102.5 °F or greater, the dose was 10 mg/kg.

Both its status as a prescription medication and the two-tier dosing schedule predicted that ibuprofen would be used for more severe illness than acetaminophen. This prediction was supported by a survey of 108 physicians (61 pediatricians, 47 family practitioners) conducted in 1992 [32]. More than half of the physicians in the study reported that they treated children with ibuprofen after acetaminophen failed, but none reported using acetaminophen only when ibuprofen was not effective. Further, both the minimum age and temperature at which the physicians recommended using these drugs were higher for ibuprofen than acetaminophen. It seemed clear that pediatric ibuprofen would be most commonly used among children whose illness was relatively severe or whose fever was particularly high or unresponsive to acetaminophen. Because of the greater severity of illness (and potential exposure to antibiotics or other medications), there was a reasonable basis to believe that ibuprofen users would experience relatively high rates of adverse clinical events, *unrelated to the ibuprofen itself*. Thus, to provide a valid assessment of the risks of pediatric ibuprofen, the study must be able to distinguish the risks of the drug from the risks associated with the illness for which ibuprofen was given.

To address this question, an LST design was used to conduct the Boston University Fever Study, an office-based study of the safety of ibuprofen use in children. The methods and results have been described in detail but are briefly summarized here [32,40]. The study was a practitioner-based trial designed to compare the risk of rare but serious adverse events among children treated for fever with ibuprofen or acetaminophen. Study participants were children between 6 months and 12 years of age with a febrile illness, which in the opinion of the managing physician warranted treatment with an antipyretic. Eligible children weighed at least 7 kg, had no contraindication to receiving either ibuprofen or acetaminophen suspension, and were in the care of a parent who could read and follow instructions written in English. Participants were identified by community-based pediatricians and family physicians and were randomized to one of three treatment groups: acetaminophen (12 mg/kg per dose), ibuprofen (5 mg/kg per dose) or ibuprofen (10 mg/kg per dose). Both the healthcare providers and parents were blind with respect to the drug/dose administered. The primary study endpoints were hospitalization for GI bleeding, acute renal failure, anaphylaxis, or Reye syndrome occurring within four weeks of enrollment. Follow-up data collection was conducted by mailed questionnaire, supplemented with telephone interviews and review of medical records for hospital admissions.

A total of 84 142 children were enrolled by 1735 practitioners. Approximately 28 000 children were randomized to each of the three treatment arms; demographic and clinical characteristics of the participants were balanced in the three groups. Overall, 795 children were hospitalized for any reason during follow-up; the risk of hospitalization did not vary significantly by treatment group (range 0.89–0.99%). Four children were hospitalized for GI bleeds, all of whom had been assigned to receive ibuprofen – two in each dose group. The risk of hospitalization for GI bleeding among all ibuprofen-treated children, 7.2 per 100 000 (95% confidence interval [CI] 2–18 per 100 000), was

not significantly greater than the risk among children randomized to acetaminophen, 0 per 100 000 (95% CI 0–11 per 100 000). No children were hospitalized for any of the other primary study endpoints, and the risk of each was no more than 5.4 per 100 000 (based on the upper bound of the 95% CI). The study provided substantial evidence of the safety of short-term use of ibuprofen in children and was used to support an application for OTC sales of pediatric ibuprofen suspension in the US.

Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) LST

Ziprasidone is an atypical antipsychotic for the treatment of patients with schizophrenia. Early evidence from premarketing clinical trials indicated that ziprasidone might be associated with improved lipids and lower incidence of weight gain. Despite the need for additional effective treatments for schizophrenia and these potential benefits, concerns about its safety, specifically whether the modest QTc prolongation associated with the drug would translate into increased mortality in patients using it in the real world, were voiced at the time of ziprasidone's review and approval by regulatory authorities in the US and Europe. QTc prolongation is of clinical concern because of its potential to induce torsade de pointes and other serious ventricular arrhythmias resulting in sudden death.

Prior to approval, the Sponsor (Pfizer, Inc.) completed a clinical study comparing six antipsychotics. The study found that the mean QTc prolongation was approximately 9–14 milliseconds greater for ziprasidone than for several others tested but approximately 14 milliseconds lower than thioridazine (a drug associated with reports of sudden death that resulted in a black box warning in the US). The study found similar results when the drugs were administered in the presence of a metabolic inhibitor. Although drugs associated with the risk of a greater degree of QTc prolongation than ziprasidone had been shown to increase the risk of sudden death, the precise relationship between QTc prolongation and the risk of serious adverse cardiac events was unknown at the time of ziprasidone's approval [41,42].

It is in this context that an LST was selected as the most appropriate study design for postmarketing safety evaluation. In typical psychiatric practice, patients treated with a new medication may be systematically different from those treated with other drugs, due to prescribers' channeling of the drug to patients with more severe schizophrenia and/or co-morbidities and risk factors. This possibility existed because ziprasidone was the newest antipsychotic at that time, and most likely to be used in patients who had failed prior therapies. In addition, despite the potential risk associated with QTc prolongation, it was possible that patients treated with ziprasidone might differ from those treated with other antipsychotic drugs, due to prescribers' channeling of the drug to patients with underlying cardiovascular disease or metabolic illnesses, especially given the low propensity for weight gain associated with ziprasidone. Given these likely selection phenomena, random allocation of patients was the only approach providing the certainty of an unbiased comparison between groups.

ZODIAC, a regulatory requirement for Pfizer from both the US FDA and the Swedish Medicinal Products Agency (MPA), compared the safety of ziprasidone and olanzapine. The methods and results have been described in detail but are briefly summarized here [41–44]. Olanzapine was chosen as the medication for the comparison group since it is also an atypical antipsychotic medication without the same effect as ziprasidone on the QTc prolongation. The study, which intended to randomize 18 000 patients, was unprecedented in psychiatric research, both in size and design. The primary objectives of the study were to estimate relative all-cause, nonsuicide, suicide, cardiovascular

and sudden death mortality among users of ziprasidone and olanzapine. The secondary objectives were to estimate the relative incidence of all-cause hospitalization and hospitalization for arrhythmia, myocardial infarction, or diabetic ketoacidosis, and to determine the rate of treatment discontinuation. The study design was not expected to permit a statistically powered evaluation of the differences in the incidence of uncommon but important adverse effects, such as torsade de pointes and sudden death. Instead, nonsuicide mortality was chosen as the primary endpoint for which the study was powered since even a larger increase in an uncommon cause of death like torsade de pointes or sudden death could be counterbalanced by a small decrease in a more common cause of death, like atherosclerotic events. In addition, torsade de points would not be reliably detected as part of normal medical or psychiatric care because of its rarity and the absence of frequent (or any) ECG testing in the routine clinical settings in which the study was carried out. The all-cause nonsuicide mortality aggregate measure was therefore deemed to be the most important and appropriate primary outcome measure.

After the enrolling physician determined a patient's eligibility and obtained informed consent, brief information, including demographics, disease severity, cardiac risk factors, and prior antipsychotic medication use, was collected on a baseline questionnaire. Social security numbers and information on up to two alternate contact/family caregivers were also collected. Following random assignment of medication, no further study-related interventions, tests, or visits were required. Physicians and patients were allowed to change regimens and dosing of the assigned study medication, and concomitant medications were permitted. In the US, patients were prescribed their assigned medicine and provided a pharmacy card to cover the costs of the medicine at a local pharmacy. Patients were followed as clinically appropriate and outcomes assessed for up to one year, regardless of how long they used the assigned medication. Information on the patients' vital status and whether or not they were hospitalized was obtained through follow-up with the treating physician or other designated member of the medical care team, family caregiver or through national death indices. A scientific steering committee was responsible for oversight of the study, a data safety monitoring board for safeguarding study participants, and an endpoint committee for assessing whether reported events met study endpoint criteria without knowledge of the assigned treatment.

ZODIAC enrolled 18 154 patients between February 2002 and February 2006 from 18 countries, including Argentina, Brazil, Chile, Hong Kong, Hungary, Korea, Malaysia, Mexico, Peru, Poland, Romania, Singapore, Slovakia, Sweden, Taiwan, Thailand, Uruguay, and the US. The primary analyses found no difference between the ziprasidone and olanzapine treatment arms with respect to nonsuicide mortality (relative risk (RR) 1.02, 95% CI 0.79–1.39). The incidence of nonsuicide mortality within one year of initiating medication was 0.91 for ziprasidone (n = 9077) and 0.90 for olanzapine (n = 9077). This finding was confirmed in numerous sensitivity analyses. The risk of all-cause hospitalization was 39% higher among persons randomized to ziprasidone versus olanzapine (RR 1.39, 95% CI 1.29–1.50). Analyses of the remaining secondary outcomes indicated no difference between the ziprasidone group and the olanzapine group. These findings were also supported in numerous *post hoc* analyses requested by the FDA [43].

In summary, despite the known risk of QTc prolongation with ziprasidone treatment, the findings of this study failed to show that ziprasidone is associated with an elevated risk of nonsuicidal mortality relative to olanzapine in real-world use and were incorporated into the approved medicine label in the United States and Europe.

Large Simple Trials Using Routine, Electronic Data Capture in Healthcare Systems

Data routinely collected within the healthcare system can sometimes be used to improve efficiency of LSTs. This can include data from electronic health records (EHRs) with information recorded by clinical staff at the point of care (e.g., in hospitals or outpatient clinics), administrative claims data (e.g., the Veterans Affairs database in the US), national/regional registries (e.g., population-wide databases in Sweden and Denmark), and patient disease/condition/drug registries (e.g., CORONNA rheumatoid arthritis registry).

Electronic health records and registries have been used in conventional RCTs to collect long-term follow-up data. An example is the West of Scotland Coronary Prevention Study which was a primary prevention trial in 45–64-year-old men with high low-density lipoprotein cholesterol. Over 6000 men were randomized to receive pravastatin once daily or placebo for an average of 4.9 years. Subsequent linkage to EHRs and registries allowed follow-up of major cardiovascular events for over 20 years [45]. Another example is the Scandinavian Simvastatin Survival Study that used national cancer registries to determine cancer endpoints five years after closure of the trial [46].

The use of EHRs and registries for the identification of potential trial participants has been increasing in recent years (see Chapter 13). As an example, the EHR4CR project has developed a platform that utilizes data from hospital EHR systems for clinical trials feasibility assessment and patient recruitment. The platform can connect securely to the data within multiple hospital EHR systems to assess the feasibility of a trial and locate the most relevant hospital sites [47].

The most recent development is the use of EHRs and registries for identification and recruitment as well as follow-up for trials, providing an efficient and reusable infrastructure for LSTs [48,49]. An example of this type of trial is the Retropro trial which recruited patients with high cardiovascular risk in routine clinical care and randomized them between two licensed medications (simvastatin and atorvastatin) [50]. The UK Clinical Practice Research Datalink® (CPRD®) EHR database was used to preselect potentially eligible patients (i.e., high cardiovascular risk estimated by a risk score). The CPRD includes anonymized EHRs for over 5 million patients currently registered at a participating general practice and can be linked to other datasets such as the national registry of hospital admissions, death certificates, and disease registries (see Chapter 13) [51]. As infrastructure for large simple trials, the CPRD can be used to identify eligible subjects who then can be consented and enrolled by their general practitioners. CPRD data are anonymized prior to being made available to the study research team (which uses coded patient IDs). In the Retropro trial, for example, clinicians were able to recruit preselected patients during consultation or during dedicated appointments. Software within the EHR system reminded clinicians through a system alert when a patient eligible for inclusion in the trial consulted the clinician. The flagging software also triggered a link to the study website, which randomized trial participants after guiding the clinician to confirm eligibility and obtain informed consent. The data from the EHRs and linked registries were used to collect follow-up outcomes (including persistence to treatment and major clinical outcomes). Suspected adverse drug reactions were also reviewed by searching the side effect fields or main clinical fields of the EHR. The quality of outcome assessment was strengthened by implementing the prospective randomized open blinded endpoint (PROBE) assessment with case adjudication by clinicians masked to treatment allocation [52].

Another example of an embedded LST is the insulin trial within the US Veteran Affairs

Healthcare System (VA). This trial compared two methods (a sliding scale versus a weight-based regimen) for determining the dose of subcutaneously administered insulin to be used in hospitalized patients, and was carried out to test the feasibility of conducting what the investigators refer to as a point-of-care clinical trial (POCCT). To accommodate this pilot trial, the VA data collection system was modified to support enrollment, randomization component, and better capture of study endpoints [53]. Order-entry screens at three VA hospitals were modified to include an option to enroll participants into the trial comparing these two regimens. Election of the menu choice "no preference for insulin regimen" triggered the EHR workflow to notify the research nurse to obtain informed consent, automatically place a note of participation in research in the medical record, and randomly assign treatment. The primary outcome of this ongoing trial was the length of hospital stay, and secondary outcomes included measures of glycemic control, all of which were ascertained from the EHR database.

The ADAPTABLE Aspirin Study is an ongoing three-year pragmatic trial comparing the effectiveness of low- and high-dose aspirin to prevent myocardial infarction and stroke in patients with cardiovascular disease. This study is embedded in PCORnet, the National Patient Centered Clinical Research Network, which is a distributed health data research network with EHRs and administrative data [54] (see Chapter 25).

Large simple trials embedded in health systems or registries may use cluster randomization, an alternative to trials that randomize individual patients. In cluster trials, entire areas or health service organizational units are randomly allocated to intervention or control groups with outcomes evaluated for individuals within each cluster. Cluster trials are increasingly used in public health and health services research and are especially important in the evaluation of health service and public health interventions [55]. An example is a cluster trial that evaluated whether antibiotic prescribing for respiratory tract infections can be reduced by reminding clinicians of the recommended clinical practice guidelines by providing an alert in the EHR during consultation [56].

Embedded LSTs with randomization to medicines are still relatively uncommon; however, healthcare systems and research groups are developing these capabilities, as noted above. The greatest opportunities are currently for research questions that help drive a learning healthcare system [57] such as the examples above in which the study results are directly relevant to patient care; the results help providers and payers identify regimens that maximize patient benefit and safety. As more of these studies reach the peer-reviewed literature, we suspect more healthcare systems will recognize the potential benefits and efficiencies, and will develop the capability to conduct embedded point-of-care LSTs.

The Future

With accelerated approval of new medications and rapid increases in their use, there may be a greater need for large postmarketing studies capable of randomizing exposures in order to assess small differences in risk. In the absence of techniques that reliably control for confounding by indication in observational studies, there may be a growing need for LSTs to evaluate larger relative risks. By virtue of minimal restrictions on participant eligibility, LSTs are more likely than classic randomized clinical trials to reflect the true benefits and safety of medications when used in actual clinical practice. The generalizability of the results of LSTs and other pragmatic clinical trials makes these studies particularly valuable to regulators and policy makers and may lead to increased use of these studies.

It is clear that safety LSTs can be conducted. It remains less clear, however, how frequently the factors that support the need for a very large trial (see Box 32.1) will converge with those that permit such a trial to be carried out (see Box 32.2). A review published in 2011 [58] discovered that only 13 LSTs evaluating safety outcomes as a primary endpoint had been completed or were ongoing prior to 2010. Among those identified, earlier studies tended to use more elements associated with a controlled clinical trial (e.g., double-blinds) whereas more recent studies had the most characteristics associated with observational epidemiology. Nonetheless, all the LSTs identified utilized observational methods of follow-up, but did so with varying degrees of intervention to meet the study objectives, primarily in the form of scheduled visits or required laboratory and diagnostic tests.

Although the LST design may be most appropriate when studying "hard" outcomes such as death or hospitalization, the same review found that researchers used the design to compare "soft" outcomes (e.g., incidence of physician-reported symptomatic hypotension and patient-reported complaints) or outcomes that require regular measurement (e.g., pulmonary function). This finding was unexpected since much of the literature on the use of LSTs for examining the therapeutic or preventive effects of an intervention suggest they are best suited to studying "hard" outcomes to avoid assessment and reporting bias. This may also be indicative of researchers' need to measure softer outcomes that are associated with and predictive of hard outcomes that occur later (e.g., death, cancer). Other unexpected differences noted were the use of person-time on treatment as a secondary analysis, which results in potentially biased comparisons between treatment groups (i.e., because participants discontinue their medications for reasons that may be related to potential outcomes), and sample sizes that were smaller than thousands of patients.

Large simple trials are likely rare as a result of the operational, financial, and scientific hurdles of implementing the design. Substantial resources are required to accrue large sample sizes; collect multiple forms of outcome data; manage hundreds of participating investigators and sites; and ensure appropriate scientific and ethical oversight. The lack of research infrastructure for conducting research at sites or with physicians inexperienced with randomized trials, including the complex regulations governing interventional study implementation and prohibitive financial costs, are barriers to conducting LSTs. This complexity illustrates that while the design is intended to be "simple" for participating healthcare providers and patients, it might be more accurately described as a "simplified" trial. Financial considerations are also important. Twelve of the 13 LSTs identified in the review described earlier were funded by the pharmaceutical or consumer products industry, six of which were postapproval commitments to the FDA or an EU regulatory agency [58].

A simple testable safety hypothesis, a motivated patient and physician population, and the ability to follow up patients to assess outcomes are important for the feasibility of a safety LST. Clearly, evaluating differences in very rare (e.g., Stevens–Johnson syndrome) or long-latency (e.g., cancer induction) outcomes is generally not practical with the LST design. Even hard outcomes such as death are difficult to study in a randomized study if the incidence of the outcome is low or the observed rate in the study is lower than expected when initial sample size calculations are performed. Because the design uses randomization, an appropriate comparator acceptable to physicians and patients is also a critical factor in the success of an LST. Equipoise must exist, and if a new medication or vaccine has a real or perceived benefit over available treatments, the LST design will not be feasible. The lack of equipoise may also become a challenge in situations where the study enrollment period is very lengthy. As knowledge of the

benefits and risks of treatments evolves, leading to less clinical uncertainty about the type of patients that might benefit from a particular study medication, investigators may find it increasingly difficult to enroll study participants.

Large simple trials should be considered as a potential design to understand the safety of any health intervention. However, the design's use with novel medicines, vaccines, diagnostics, and biologics is likely to be limited. The feature of randomization alone limits its application, since in some disease areas there may be no logical or ethical comparator (i.e., first-in-class medicine or vaccine, rare diseases, oncology). Biopharmaceutical development is increasingly focused on smaller indicated patient populations. This will result in insufficient exposure to reach the required event rate or, when feasible, it may take too long to accrue exposure to be acceptable to decision makers. Even in situations where the prevalence of a condition is much greater (e.g., schizophrenia or asthma), it may take many years to accrue the necessary patient population, which to some may be too long to address important safety questions. In settings where there is no appropriate control treatment and it is not ethical to randomize between active drug and placebo, an alternative to randomizing to comparator medicines may be to randomize to different doses, when possible, and search for a dose–response relationship. Another option may be usual care as a comparator or to randomize to treatment strategies rather than specific medications.

Despite these challenges to using the LST for comparative safety evaluation, the design is particularly suited to research questions in which confounding by indication or severity is likely to be pronounced and difficult to measure or control, but where assessment under routine care is important for decision making. LST designs are similar to observational studies in that they can, in principle, be effectively used to study the safety of health interventions in patient populations not typically exposed in clinical trials, such as the elderly, very young, or those with multiple comorbidities, and understand the safety of a health intervention as it is used with multiple concomitant prescriptions or OTC medications under routine medical care. With the increasing demand for real-world evidence to guide public health, regulatory, and clinical decisions, pharmacoepidemiologists are well suited to design and conduct LSTs and, most importantly, to develop more efficient methods of participant selection and follow-up data collection that can make these studies a more common option for the evaluation of small but important risks of medication use.

References

1 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2004; **58**: 323–37.

2 Sobel RE, Reynolds RF. Integrating evidence from multiple sources to evaluate post-approval safety: an example of sildenafil citrate and cardiovascular events. *Curr Med Res Opin* 2008; **24**(7): 1861–8.

3 Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. *JAMA* 1999; **281**(9): 824–9.

4 Kaitin KI. The new drug approvals of 1999, 2000, and 2001: drug development trends a decade after passage of the Prescription Drug User Fee Act of 1992. *Drug Inf J* 2003; **37**: 357–71.

5 Tufts Center for the Study of Drug Development. *Tufts Impact Report*. Boston: Tufts Center for the Study of Drug Development, 2008.

6 US Food and Drug Administration. *Report to Congress: Report on Postmarketing Studies*. Washington DC: US Food and Drug Administration; 2002.

7 Csizmadi I, Collet JP, Boivin JF. Bias and confounding in pharmacoepidemiology. In: Strom B, ed. *Pharmacoepidemiology*, 4th edn. Chichester: John Wiley & Sons, 2005, pp. 791–809.

8 D'Agostino RB Jr, d'Agostino RB Sr. Estimating treatment effects using observational data. *JAMA* 2007; **297**(3): 314–16.

9 Walker A. Confounding by indication. *Epidemiology* 1996; **7**: 335–6.

10 Joffe MM. Confounding by indication: the case of calcium channel blockers. *Pharmacoepidemiol Drug Saf* 2009; **9**(1): 37–41.

11 Kizer JR, Kimmel SE. Epidemiologic review of the calcium channel blocker drugs. *Arch Intern Med* 2001; **161**: 1145–58.

12 Ahdieh L, Gange SJ, Greenblatt R, *et al*. Selection by indication of potent antiretroviral therapy use in a large cohort of women infected with human immunodeficiency virus. *Am J Epidemiol* 2000; **152**(10): 923–33.

13 Garrett J, Lanes S, Kolbe J. Risk of severe life threatening asthma and beta agonist type: an example of confounding by severity. *Thorax* 1996; **51**: 1093–9.

14 Dunn N, White I, Freemantle S, Mann R. The role of prescribing and referral bias in studies of the association between third generation oral contraceptives and increased risk of thromboembolism. *Pharmacoepidemiol Drug Saf* 1998; **7**: 3–14.

15 Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of β_2-agonists. *Am J Epidemiol* 1996; **144**(12): 1161–9.

16 LaPorte J, Ibanez L, Vidal X, *et al*. Upper gastrointestinal bleeding associated with the use of NSAIDs: newer versus older agents. *Drug Saf* 2004; **27**: 411–20.

17 Mamdani M, Rochon P, Juurlink D, *et al*. Observational study of upper gastrointestinal haemorrhage in elderly patients given selective cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs. *BMJ* 2002; **325**: 624.

18 Vandenbroucke JP, Helmerhorst FM, Bloemenkamp K, *et al*. Third-generation oral contraceptive and deep venous thrombosis: from epidemiologic controversy to new insight in coagulation. *Am J Obstet Gynecol* 1997; **177**: 887–91.

19 Walker AM. Newer oral contraceptives and the risk of venous thromboembolism. *Contraception* 1998; **57**: 169–81.

20 Wolfe F, Flowers N, Burke T, Arguelles L, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. *J Rheumatol* 2002; **29**: 1015–22.

21 Avorn J. In defense of pharmacoepidemiology – embracing the ying and yang of drug research. *N Engl J Med* 2007; **357**(22): 2219–21.

22 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**: 409–20.

23 Buring JE, Hennekens CH. The contribution of large, simple trials to prevention research. *Prev Med* 1994; **23**: 595–8.

24 Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995; **48**(1): 23–40.

25 ISIS-1 Group. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet* 1986; **2**: 57–66.

26 ISIS-2 Group. Randomised trial of intravenous streptokinase, oral, aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; **2**: 349–60.

27 ISIS-3 Group. A randomised comparison streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41,299 cases of suspected acute myocardial infarction. *Lancet* 1992; **339**: 753–70.

28 ISIS-4 Group. A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995; **345**: 669–82.

29 Steering Committee of the Physicians Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Heath Study. *N Engl J Med* 1989; **321**(3): 129–35.

30 Smart Study Group. CD4+ count-guided interruption of antiretroviral treatment. *N Engl J Med* 2006; **355**(22): 2283–96.

31 Hasford J. Drug risk assessment: a case for large trials with lean protocols. *Pharmacoepidemiol Drug Saf* 1994; **3**: 321–7,.

32 Mitchell AA, Lesko SM. When a randomized controlled trial is needed to assess drug safety: the case of paediatric ibuprofen. *Drug Saf* 1995; **13**: 15–24.

33 US Food and Drug Administration. *Guidance for Industry: Premarketing Risk Assessment*. Washington DC: US Food and Drug Administration, 2005.

34 European Medicines Agency. *ENCePP Guidelines on Methodological Standards in Pharmacoepidemiology*, 6th edn. London: European Medicines Agency, 2017.

35 Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. A policy perspective. *Ann Intern Med* 1979; **90**: 257–61.

36 Gray R, Clarke M, Collins R, Peto R. Making randomized trials larger: a simple solution? *Eur J Surg Oncol* 1995; **2**: 137–9.

37 Yusuf S. Reduced mortality and morbidity with the use of angiotensin-converting enzyme inhibitors in patients with left ventricular dysfunction and congestive heart failure. *Herz* 1993; **18**(Suppl): 444–8.

38 Hennekens CH, Buring JE. Methodologic considerations in the design and conduct of randomized trials: the U.S. Physicians' Health Study. *Control Clin Trials* 1989; **10**: 142S – 50S.

39 Lee IM, Cook NR, Manson JE, Buring JE, Hennekens CH. Beta-carotene supplementation and incidence of cancer and cardiovascular disease: the Women's Health Study. *J Natl Cancer Ins*. 1999; **91**(24): 2102–6.

40 Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen: a practitioner-based randomized clinical trial. *JAMA* 1995; **273**: 929–33.

41 Strom BL, Faich GA, Reynolds RF, *et al*. The ziprasidone observational study of cardiac outcomes (ZODIAC): design and baseline characteristics. *J Clin Psychiatry* 2008; **69**: 114–121.

42 Strom BL, Eng SM, Faich G, *et al*. Comparative mortality associated with ziprasidone and olanzapine in real-world use among 18,154 patients with schizophrenia: the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC). *Am J Psychiatry* 2010; **168**: 193–201.

43 Geier JL, Karayal ON, Lewis M, *et al*. Methodological challenges in the coding and adjudication of sudden deaths in a large simple trial with observational follow-up: the ziprasidone observational study of cardiac outcomes (ZODIAC). *Pharmacoepidemiol Drug Saf* 2011; **20**(11): 1192–8.

44 Kolitsopoulos FK, Strom BL, Faich G, *et al*. Lessons learned in the conduct of a global, large simple trial of treatments indicated for schizophrenia. *Cont Clin Trials* 2013; **34**: 239–47.

45 Ford I, Murray H, McCowan C, Packard CJ. Long-term safety and efficacy of lowering low-density lipoprotein cholesterol with statin therapy: 20-year follow-up of West of Scotland

Coronary Prevention Study. *Circulation* 2016; **133**(11): 1073–80.

46 Strandberg TE, Pyörälä K, Cook TJ, *et al.* Mortality and incidence of cancer during 10-year follow-up of the Scandinavian Simvastatin Survival Study (4S). *Lancet* 2004; **364**(9436): 771–7.

47 Electronic Health Records for Clinical Research. www.ehr4cr.eu (accessed April 22, 2019).

48 Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. *N Engl J Med* 2016; **374**: 2152–8.

49 van Staa TP, Goldacre B, Guilliford M, *et al.* Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ (Online)* 2012; **344**: 7843.

50 van Staa TP, Dyson L, McCann G, *et al.* The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014; **18**(43): 1–146.

51 Herrett E, Gallagher AM, Bhaskarane K, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**(3): 827–36.

52 Hansson L, Hedner T, Dahlöf B. A novel design for intervention trials: prospective randomized open blinded end-point. *Blood Pressure* 1992; **1**(2): 113–19.

53 D'Avolio L, Ferguson R, Goryachev S, *et al.* Implementation of the department of Veterans Affairs' first point-of-care clinical trial. *JAMA* 2012; **19**(e1): e170–6.

54 National Patient Centered Network for Clinical Outcomes. ADAPTABLE: The Aspirin Study. http://theaspirinstudy.org/ (accessed April 22, 2019).

55 Mallick U, Bakhai A, Wang D, Flather M. Cluster randomised trials. In: Wang D, ed. *Clinical Trials*. London: Remedica, 2006, pp. 141–51.

56 Gulliford MC, van Staa T, Dregan A, *et al.* Electronic health records for intervention research: a cluster randomized trial to reduce antibiotic prescribing in primary care (eCRT Study). *Ann Fam Med* 2014; **12**(4): 344–51.

57 Choudhry NK. Randomized, controlled trials in health insurance systems. *N Engl J Med* 2017; **377**: 957–64.

58 Reynolds RF, Lem JA, Gatto N, Eng S. Is the large simple trial design used for comparative, post-approval safety research? A review of a clinical trials registry and the published literature. *Drug Saf* 2011; **34**: 799–820.

33

The Use of Pharmacoepidemiology to Study Beneficial Drug Effects

Brian L. Strom[1] and the late Kenneth L. Melmon[2]

[1] *Rutgers Biomedical and Health Sciences, Newark, NJ, USA*
[2] *Formerly Stanford University School of Medicine, Stanford, CA, USA*

In order to be approved for marketing in the United States, drugs must be proven to be safe and effective using "adequate and well-controlled investigations." Earlier chapters in this book have shown that this premarketing information often is insufficient to provide some of the information about drug toxicity which is clinically most important. The same applies to information about drug efficacy.

In this chapter we will begin by clarifying the different definitions of various types of beneficial drug effects. Then we will discuss the need for *postmarketing studies of drug effectiveness*. Next, we will present the unique methodologic problems raised by studies of beneficial drug effects, as well as potential solutions to these problems. Finally, we will evaluate the frequency with which these proposed solutions might be successful. Specific examples of approaches to the study of efficacy also will be presented.

Definitions

There are at least four different types of measurable drug effects of interest to a prescriber. *Unanticipated harmful effects* are the unwanted effects of drugs that could not have been predicted on the basis of their preclinical pharmacologic profile or the results of premarketing clinical studies. These effects are most often type B adverse reactions, as defined in Chapter 1. For example, chloramphenicol was not known to cause aplastic anemia at the time it was marketed [1], nor was the skeletal muscle pain associated with use of HMG-CoA reductase inhibitors. A major research challenge is to discover medically important unanticipated harmful effects as soon as possible after drug marketing. Quantitation of the incidence of these effects is medically useful as well.

Anticipated harmful effects are unwanted effects of drugs that could have been predicted on the basis of preclinical and premarketing studies. They can be either type A reactions or type B reactions (see Chapter 1). One example is the syncope that sometimes occurs after patients take their first dose of prazosin [2]. Although this effect was known to occur at the time of marketing, a major question remaining to be answered was how often the event occurred. The dominant research challenge that this type of drug effect presents is establishing its incidence.

Unanticipated beneficial effects are desirable effects of drugs that were not anticipated at the

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

time of drug marketing. Although these effects may be medically useful, they are nevertheless side effects, if they are not the purpose for which the drug was given. An example of an unanticipated beneficial effect is aspirin's ability to decrease the probability of a subsequent myocardial infarction in patients who were given the drug for its analgesic or antiinflammatory action [3]. This effect was confirmed only after aspirin had been used as an analgesic for many years. A major research challenge is to discover this type of drug effect. For example, given the similar (although not exactly the same) mechanism of nonaspirin nonsteroidal antiinflammatory drugs, these drugs may have the same beneficial effects as aspirin [4]. However, recent data suggest that, in contrast, these medications might increase the risk of cardiovascular events.

Anticipated beneficial effects are the desirable effects that are known to be caused by the drug. They represent the reason for prescribing the drug. The study of anticipated beneficial effects has three aspects.

A study of drug *efficacy* investigates whether a drug *has the ability* to bring about the intended effect. In an ideal world, with perfect compliance, no interactions with other drugs or other diseases, etc., could the drug achieve its intended effects? Drug efficacy usually is studied using a randomized clinical trial.

In contrast, a study of drug *effectiveness* investigates whether, in the real world, a drug *in fact* achieves its desired effect. For example, a drug given in experimental conditions might be able to lower blood pressure but if it causes such severe sedation that patients refuse to take it, it will not be effective. Thus, an efficacious drug may lack effectiveness. Studies of drug effectiveness usually are performed after a drug's efficacy has been established. In contrast, if a drug is demonstrated to be effective, it also is obviously efficacious. Studies of drug effectiveness generally would best be conducted using nonexperimental study designs, although clinical trials using pragmatic designs may also be utilized (see Chapter 32). However, these raise special methodologic problems, which are discussed later.

Lastly, a study of *efficiency* investigates whether a drug can bring about a *desired effect at an acceptable cost*. This type of assessment falls in the province of health economics, which is discussed in Chapter 34.

Note that the outcome variable for any of these studies can be of multiple different types. They can be clinical outcomes (diseased/undiseased) (see Chapter 37 for a discussion of the validity issues involved in measuring such outcomes); measures of patient-reported outcomes (see Chapter 42), measures of utility, such as global measures of the desirability of certain clinical outcomes (see Chapters 37 and 42); economic outcomes (see Chapter 34); etc. Regardless, the same methodologic issues apply to each.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

In order to make optimal clinical decisions about whether to use a drug, a prescriber needs to know whether, and to what degree, the drug actually is able to produce the intended effect (Box 33.1) [5]. Premarketing randomized clinical trials generally provide information on whether a drug can produce at least one beneficial effect. Specifically, premarketing studies generally investigate the efficacy of a drug relative to a placebo, where both are used to treat a particular illness. These premarketing studies of efficacy tend to be conducted in very atypical clinical settings, compared to those in which the drug ultimately will be used. Patient adherence (see Chapter 38) during these studies is typically higher than in actual practice, and the patients included are similar to each other in age and sex, do not have other diseases, and are not taking potentially interacting drugs (see Chapter 40). Such restrictions

Box 33.1 Clinically important information about intended beneficial effects of drugs

1) Can the drug have the desired effect?
2) Does the drug actually achieve the desired effects when used in practice?
3) Can and does the drug have other beneficial effects, including long-term effects for the same indication?
4) Can the drug achieve these desired effects better than other alternative drugs available for the same indication?
5) For each of the above, what is the magnitude of the effect in light of the many different factors in medical practice that might modify the effect, including:
   - variations in drug regimen: dose per unit time, distribution of dose over time, duration of regimen?
   - characteristics of the indication: severity, subcategories of the illness, changes over time?
   - characteristics of the patient: age, sex, race, genetics, geographic location, diet, nutritional status, adherence, other illnesses, drugs taken for this or other illness (including tobacco and alcohol), etc.?

Modified from Strom BL, Melmon KL, Miettinen OS. Postmarketing studies of drug efficacy: Why? *Am J Med* 1985; **78**: 475–80. Reproduced with permission of Elsevier.

maximize the ability of premarketing studies to demonstrate a drug's efficacy, if the drug actually is efficacious.

Additional information may then be needed on whether, in the world of daily medical practice, the drug actually achieves the same beneficial effects and whether it can and does have other beneficial effects. In addition, at the time of marketing there may be few data on a drug's efficacy relative to other medical or surgical alternatives available for the same indication.

Finally, a number of factors that are encountered in the practice of medicine can modify a drug's ability to achieve its beneficial effects. Included are variations in drug regimen, characteristics of the indication for the drug, and characteristics of the patient receiving the drug, including demographic factors, nutritional status, the presence of concomitant illnesses, the ingestion of drugs, and so on. Many, if not most, of these factors that can influence the effects of drugs are not fully explored prior to marketing.

In order to quantitate the need for postmarketing studies of the beneficial effects of drugs, a comparison was made of the 100 most common drug uses in 1978 (drug–indication pairs) to the information available to the FDA at the time of its regulatory decisions about the marketing and labeling of the drugs involved in these uses [5]. The comparison was restricted to drugs approved after 1962, when the Kefauver–Harris Amendments first introduced a requirement for the submission of data about drug efficacy prior to approval of a drug for marketing.

Of the 100 common drug uses, 31 had not been approved by the FDA at the time of initial marketing, and 18 still had not been approved at the time of the comparison. Eight of the 18 unapproved uses were probably medically and therapeutically inappropriate. For example, the use of antibiotics is not justified for the treatment of viral infections, but such use was common. Other unapproved drug–indication pairs could well have been quite appropriate, but the regulatory process does not need to and did not reflect the current medical practice.

Of the 100 common drug uses, eight were based on the assumption that a drug had a particular long-term effect, but only an intermediate effect had been studied prior to marketing. For example, antihypertensive drugs are used for their presumed ability to prevent long-term cardiovascular complications, but are approved for marketing on the basis of their ability to lower blood pressure. Five of the 100 common

drug uses may have been for either the intermediate effect or the long-term effect of the drugs, but only the intermediate effect was studied prior to marketing. For example, hypoglycemic agents may be used to control the symptoms of diabetes or to prevent the vascular complications of diabetes, but only the former were studied before drug marketing.

Drugs other than those in the list of 100 common uses were sometimes prescribed as treatment for each of the 52 indications included in those 100 uses. Yet, eight of the uses involved drugs whose effects relative to alternative drugs had not been studied prior to marketing.

The 100 common drug uses also included examples of clinical factors that are able to modify the effects of the drug, but these were not discovered until after drug marketing. Some are listed in Table 33.1 [6–19]. In addition, additional prescriptions accompanied 62% of the prescriptions studied, and 41% of the prescriptions were for patients who had illnesses other than just the one that the drug was being used to treat. Of the 100 common drug uses, the mean number of concomitantly administered drugs ranged from 0.04 to 2.1. The mean number of concomitant diagnoses ranged from 0.1 to 1.2. Yet, for none of the uses was the potential for modification of the drug effect by concomitant drugs or concomitant diagnoses fully explored before marketing.

The proportion of prescriptions which were for patients less than age 20 ranged from 0.0%, for 43 of the uses, to 97%. Yet many of these uses had not been tested in children prior to marketing. Analogously, only three of the drugs were approved for use in pregnant patients, yet we know that drug use in pregnancy was common, even then [20–22].

Thus, this study revealed considerable gaps in the information about beneficial drug effects at the time of drug marketing. These deficiencies in the available information should not be surprising, nor should they be considered inadequacies that ought to prevent the release of the drug to the marketplace. The data needed for clinical decisions are frequently and understandably different from those needed for regulatory decisions. Studies performed prior to marketing perforce are focused predominantly on meeting appropriate regulatory requirements, and only secondarily on providing a basis for optimal therapeutic decisions. This regulation is not aimed at telling a physician precisely how an agent should be used. The FDA is not allowed to regulate physicians but, rather, pharmaceutical manufacturers. In addition, the FDA does not initiate its own studies of drug effects, but generally evaluates those submitted to it by manufacturers.

Finally, there are reasonable logistical limitations on what can be expected prior to marketing, without undue cost in time and resources, as well as delaying the availability of a chemical entity with a proven potential for efficacy. Thus, it seems that more studies of beneficial drug effects are needed, perhaps as a routine part of postmarketing drug surveillance.

Methodologic Problems To Be Addressed by Pharmacoepidemiologic Research

Chapter 3 introduced the concept of a confounding variable, which is a variable other than the risk factor and outcome variable under study which is related independently to each of the other two and, thereby, can create an apparent association or mask a real one. This is discussed in more depth in Chapter 43. Studies of intended drug effects present a special methodologic problem of confounding by the indication for therapy [23,24]. In this case, the risk factor under study is the drug being evaluated and the outcome variable is the clinical condition that the drug is supposed to change (cure, ameliorate, or prevent). In clinical practice, one would expect

Table 33.1 Examples of factors determining drug efficacy that were demonstrated after marketing, selected from the 100 most common drug uses of 1978.

Factors	Drug	Indication	Comments	Reference
Regimen				
Dose per unit time	Ibuprofen	Rheumatoid arthritis, osteoarthritis	Daily dosage initially approved proved to be suboptimal	5
Distribution of dose over time	Furosemide	Congestive heart failure	Efficacy improved by more frequent, smaller doses	6
Duration	Clonidine	Hypertension	Tolerance develops in the absence of a diuretic	7
	Hypoglycemics (e.g., acetohexamide and tolazamide)	Diabetes mellitus	Tolerance develops in many patients	8
Indication				
Severity	Metaproterenol	Asthma	Patients with severe illness do not have a response without additional, supplementary therapy	9
Subcategories	Desipramine	Depression	May vary with endogenous versus exogenous depression	10
Changes over time	Ampicillin	Otitis media	No longer the drug of choice in some geographic areas due to bacterial resistance	11, 12
Patient				
Age	Diazepam	Anxiety	A given regimen is more effective in the aged than in the young	13
			Metabolism varies markedly from premature infants (half-life 54 hours), to full-term infants, to older children (half-life 18 hours); young children can have paradoxic reactions	14
Other illness	Gentamicin	Infection	Lower doses required in renal failure	15
Other				
Drugs	Lithium	Manic-depressive illness	Clearance impaired by diuretics, e.g., furosemide	16
	Acetohexamide	Diabetes mellitus	Many drugs interfere, by causing hyperglycemia (e.g., diuretics), displacing drug from binding sites (e.g., nonsteroidal anti inflammatory drugs), etc.	17
Diet	Diuretics (e.g., metolazone, furosemide)	Hypertension	A decrease in sodium intake can improve efficacy	18
	Lithium	Manic-depressive illness	Significant sodium depletion or excess can modify renal excretion	16

Source: Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: Why? *Am J Med* 1985; **78**: 475–80 (5).

treated patients to differ from untreated patients, as the former have an indication for the treatment. To the extent that the indication is related to the outcome variable as well, the indication can function as a confounding variable.

For example, if one wanted to evaluate the effectiveness of a beta-blocker used after a myocardial infarction in preventing a recurrent myocardial infarction, one might conduct a cohort study comparing patients who were treated with the beta-blocker as part of their usual postmyocardial infarction medical care to patients who were not treated, measuring the incidence of subsequent myocardial infarction in both groups. However, patients with angina, arrhythmias, and hypertension, all indications for beta-blocker therapy, are at increased risk of subsequent myocardial infarction. As such, one might well observe an increase in the risk of myocardial infarction, rather than the expected decrease. Thus, even if use of the drug was beneficial, it might appear to be harmful!

Confounding by the indication for the treatment generally is not a problem if a study is focusing on unexpected drug effects, or side effects, unrelated to the indication for the drug, whether they are harmful or beneficial effects. In this situation, the indication for treatment is not usually related to the outcome variable under study. For example, in a study of gastrointestinal bleeding from nonsteroidal antiinflammatory drugs (NSAIDs), the possible indications for treatment, such as arthritis, dysmenorrhea, and acute pain, have little or no relationship in and of themselves to the risk of gastrointestinal bleeding [25]. Nevertheless, sometimes the problem of confounding by indication can emerge even in studies of unexpected drug effects (beneficial or harmful). For instance, in a study of hypersensitivity reactions associated with the use of NSAIDs, the increased risk of hypersensitivity reactions evident in patients taking NSAIDs was higher in those using the drugs for acute pain than in those using the drugs for osteoarthritis and other chronic conditions. This probably was because of the intermittent ingestion of the drug by those receiving it for acute pain [26].

Although confounding by the indication is a less common problem for studies of side effects, this is not the case for studies of anticipated beneficial effects. In these studies, one would expect the indication to be more closely related to the outcome variable. In fact, the problem presented by confounding by the indication has been thought by some to invalidate nonexperimental approaches to studies of the beneficial effects of drugs. Some have felt that questions of beneficial drug effects can be addressed only by using randomized clinical trials [27]. Yet, although postmarketing randomized clinical trials certainly can be very useful, they are vexed by many of the same logistical problems, ethical restrictions, and artificial medical settings found in premarketing clinical trials.

Currently Available Solutions

Not all studies of beneficial drug effects need be randomized clinical trials (Table 33.2) [23]. First, some questions do not require any comparative (analytic) research for their answer. For these, simple clinical observations, as reported in a case report or case series, can be sufficient. For example, the efficacy and effectiveness of naloxone, used as a narcotic antagonist, is demonstrable simply through the observation of a single patient. Consider a patient comatose from an overdose of methadone. An injection of naloxone results in his prompt awakening. However, 30 minutes later, as the effects of the narcotic antagonist wear off, the patient returns to coma. Another injection of the naloxone results in awakening once more, and then later the coma returns again. This sequence of events represents a convincing demonstration of the drug's ability to have its desired effect. No elaborate studies are needed to make this point. The same would be

Table 33.2 Classification of research questions according to their problems of confounding by the indication for therapy.

Situation	Example
1. Comparative studies unnecessary	
(a) Drug effect obvious in the individual patient, or	Naloxone used for methadone overdose
(b) Drug effect obvious in a series of patients	Penicillin used for pneumococcal pneumonia
2. Confounding by the indication nonexistent: there is no indication	Measles vaccine given routinely to healthy infants
3. Confounding by the indication exists but is controllable	
(a) The indication is dichotomous	
(i) Gradations in the indication do not exist, or	Anti-Rh (D) immune globulin given to Rh (D)-negative mothers who deliver Rh (D)-positive newborns to prevent future erythroblastosis fetalis
(ii) Gradations in the indication are unrelated to the choice of treatment, or	Penicillin used for endocarditis prophylaxis in patients with congenital aortic stenosis who are undergoing tooth extraction
(iii) Gradations in the indication are unrelated to expected outcome, or	Penicillin used to prevent tertiary syphilis, given to patients with an asymptomatic positive serologic test for syphilis
(iv) Special clinical settings	Anticoagulants used after myocardial infarctions to prevent death
(b) The indication is sufficiently characterizable	Isoniazid used for tuberculosis prophylaxis in a patient with an asymptomatic positive purified protein derivative
(i) Complete characterization of the indication as it relates to choice of therapy or as it relates to expected outcome, and	
(ii) Characterization must continue after initiation of therapy	
4. Confounding by the indication exists and is not controllable	Ampicillin used to treat urinary tract infection

Source: Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? *Clin Pharmacol Ther* 1983; **34**: 1–7. Reproduced with permission of John Wiley & Sons.

true for a case series of patients treated with penicillin to treat pneumococcal pneumonia.

However, in applying this simple approach of clinical observations based on a case report or case series, the course of a patient's disease must be sufficiently predictable that one can differentiate a true drug effect from spontaneous improvement. In particular, one must be able to exclude *regression to the mean* as the mechanism of the observed change: individuals selected to participate in a study based upon the severity of their disease spontaneously and usually will tend to improve. One example would be a patient with recurrent headaches. The patient would most likely seek medical attention when the headaches are most severe or most frequent.

A spontaneous return to the baseline pattern of headaches generally could be expected. However, if the patient were treated in the interim, then the treating physician likely would view the return to normality as evidence of successful therapy, no matter what treatment was used or whether it contributed anything to the recovery.

Second, some questions about beneficial drug effects can be answered using formal nonexperimental studies, because there is no confounding by the indication. If the decision about whether to treat is not based on a formal indication, but on some other factor that may not be related to the outcome variable under study, such as the limited availability of the drug in question, then there is no opportunity for confounding by the indication. This situation occurs most commonly in studies of primary prevention. The use of measles vaccine, routinely administered to healthy infants, is one example.

Third, there are several settings in which confounding by the indication may exist but theoretically can be controlled. When the indication can be measured sufficiently well, then traditional epidemiologic techniques of exclusion, matching, stratification, and mathematical modeling can be applied. The indication clearly can be sufficiently measured if it is dichotomous or binary. In this situation, the indication either is present or absent, but has no gradations in severity. The indication also can be sufficiently measured if any gradations in severity either are unrelated to the choice of whether or not to treat or are unrelated to the expected outcome. Alternatively, sometimes one can find special clinical settings in which the gradations are not related to the choice of therapy. For example, if the availability of drugs is limited or there are consistent philosophical differences among prescribers for using or not using the drug, then gradations in the indication will not be related to the choice of therapy (although, even then, confounding could occur due to, for example, differences in patients treated among prescribers).

Finally, if an indication is graded but can be sufficiently precisely measured, it can be controlled by mathematical modeling using, for example, multiple regression. Then confounding by the indication can be controlled and ruled out as the cause for an observed beneficial effect of the drug.

More recently, researchers have used *propensity scores* towards this end [28,29]. This approach uses mathematical modeling to predict *exposure*, rather than the traditional approach of predicting *outcome* [30]. This is, essentially, a direct measure of indication. One can then use the propensity score to create categories of probability of exposure, and control for those categories in the analysis. While this approach has many attractive features, especially as a direct way to control for confounding by indication, it is important to point out that it is still dependent on identifying and measuring those variables which are the true predictors of therapeutic choice. Further, propensity scores only have advantages when there are seven or fewer outcome events per confounder [31]. When there are at least eight outcome events per confounder, logistic regression represents a preferable approach [31]. (See Chapter 43 for a more detailed discussion.)

Another relatively new approach increasingly being applied is the use of *instrumental variables*. An instrument is a variable that is causally related to the exposure of interest, only weakly related to the uncontrolled risk factors of concern, and is not itself in the causal chain. Thus, an instrument is an external factor that influences an outcome only through its effect on treatment. By controlling for the instrument, it is thought that one can control for the indication for treatment. However, finding good instruments in pharmacoepidemiology is extremely difficult. This is discussed further in Chapter 43.

When questions of intended drug effects do not fall into any of the preceding categories, *confounding by the indication cannot be controlled*. Nonexperimental study designs cannot then be

used, or they can only be used to demonstrate qualitatively some degree of beneficial effect. Specifically, if confounding by the indication is such that treated patients would have a worse clinical outcome than untreated patients, yet the outcome observed in treated patients is better than that observed in untreated patients, some degree of confidence that the drug has a beneficial effect can be built. As an example, patients treated with corticosteroids for status asthmaticus would be expected to be sicker than those not so treated. If patients receiving corticosteroids stop wheezing sooner than those not receiving corticosteroids, corticosteroids would indeed seem to have a beneficial effect. However, if the patients receiving corticosteroids do not stop wheezing sooner than those not receiving corticosteroids, the results of the study are uninterpretable. It is possible that the corticosteroids in fact have no beneficial effect but it is also possible that a beneficial effect was present but was being masked by the difference in severity between the two treatment groups.

The qualitative approach illustrated above must be used with caution. First, the effect of the confounding by indication must be opposite in direction to the expected effect of the drug. Second, the effect of the confounding by indication must be absolutely predictable in its direction. Third, the effect of the confounding by indication must be sufficiently large so as to exclude regression to the mean as an explanation for the results. Even if all of these conditions are met, the results must be interpreted only qualitatively, not quantitatively.

Examples of each of these situations are presented in Table 33.2 and discussed further in Strom *et al.* [23].

Applicability of the Proposed Approaches

How commonly are the nonexperimental approaches we have described applicable for the study of beneficial drug effects? A list of the 100 most recently approved new molecular entities as of December 1978 was studied to determine what types of nonexperimental study designs, if any, could be used to evaluate drug effectiveness [32]. After excluding from this list seven entities that were used in contact lenses, the remaining 93 drugs were examined for all potential indications and clinical outcomes that could be used to evaluate intended drug effects. Ultimately, we assessed 131 drug uses, that is 131 drug–indication pairs. Each drug use was categorized as to whether a study evaluating the effectiveness of that drug for that indication would present the problem of confounding by the indication and, if so, whether one of the approaches described above would be adequate to address it. Eighty-nine (67.9%) of the drug uses could have been evaluated using simple clinical observations, without formal comparative research. A very few of these drugs were, in fact, approved by FDA on the basis of such studies, such as nitroprusside (approved for malignant hypertension) and bretylium (approved for life-threatening arrhythmias, in patients refractory to all other antiarrhythmics). The remaining 42 drug uses required comparative research for their evaluation, because they all presented the problem of confounding by the indication. In seven of the 42 (5.3% of the total), this confounding was not an obstacle to valid nonexperimental research. Most often, the validity of the approach rested on the observation that any given physician usually used the drug to treat either all or none of his/her patients with the indication.

In the remaining 35 of the 42 uses (26.7% of the total), confounding by the indication was judged to be uncontrollable using currently available nonexperimental techniques.

To place these findings in perspective, of the 42 drug uses that required comparative research to evaluate their effectiveness, 30 could not ethically be addressed using a randomized clinical trial and a placebo control. Most of these 30 involved the use of drugs to treat infections or malignancies. In these situations, patients could

not ethically be left "untreated," that is, assigned to the placebo group.

Studies of the effects of one drug relative to another active drug, of course, gave different results. Formal comparative research was necessary for all 131 drug uses. Nonexperimental studies theoretically could be conducted validly for 94 of the 131 drug uses (71.8%). Experimental studies would be ethical for all of them.

Of course, judging theoretically that a question of effectiveness is "studiable" by a given technique is not the same as proving that a valid outcome would emerge from such a study. There are many particular details in the actual conduct of such studies that must be addressed on a case-by-case basis. It is, therefore, instructive to examine some specific examples of nonexperimental research into beneficial drug effects.

Specific Examples

Estrogens for Prevention of Osteoporotic Fractures

One of the first series of studies of drug effectiveness using rigorous nonexperimental study designs examined whether exogenous estrogens could prevent fractures in postmenopausal women with osteoporosis [33]. Biochemical studies had documented that the menopause resulted in a negative calcium and phosphorus balance, and that the balance returned toward normal with the ingestion of exogenous estrogens [34]. Studies of bone density documented that exogenous estrogens prevented the loss of bone density that was associated with the menopause [35], for as long as the estrogens were continued [36]. It seemed plausible that the use of estrogens might prevent fractures from osteoporosis, but no data directly addressed that question. On the other hand, postmenopausal estrogens had been shown to cause endometrial cancer [37,38].

A randomized clinical trial would have been the ideal way to address the effect of estrogen on fractures. However, such a study, of this prophylactic therapy, was impractical for many reasons. Although postmenopausal fractures are common, they are experienced by a sufficiently small proportion of the population during any defined time period that an extremely large sample size would be needed. Also, the study would need to be carried on for many years before a beneficial effect could begin to be seen.

Instead of a randomized clinical trial, a series of nonexperimental studies were performed. Both case–control and cohort designs were used [39–56]. In general, these studies were rigorous and well done but the question of confounding by the indication was not addressed in most of the studies [33]. In particular, most of them failed to address why some of the women received the postmenopausal exogenous estrogens and others did not. Given the data already available on the effects of estrogens on bone density [57,58] and endometrial cancer [59–62], it is reasonable to assume that some physicians might preferentially and routinely use the drugs and others might routinely avoid them [63–65]. In such a setting, nonexperimental techniques could yield valid results, unaffected by confounding by the indication (category 3.a.iv in Table 33.2). However, many physicians might try to selectively prescribe the drugs for patients who have undergone hysterectomy, because these patients are at no risk of endometrial cancer. Alternatively, some physicians may try to use the drugs only on patients who they feel are at high risk of fractures or at high risk of complications from fractures. These situations would represent uncontrollable confounding by the indication – category 4 in Table 33.2.

Finally, one might expect that the direction of the confounding by indication might be opposite to that of the drug effect, allowing one to use these data to make at least qualitative conclusions. This assumes, however, that physicians can accurately predict who is at high risk of fracture. Such a presumption was not borne out by the available data [50].

In fact, the three studies that closely examined the comparability of the study groups were able

to document that they were not comparable [39,50,52]. Specifically, one study was a case–control study within an orthopedic service, and documented that cases with fractures of the hip or radius weighed less than controls matched for age and race, had a later menopause, and more frequently were alcoholics [39]. A second was a cohort study of patients with known estrogen deficiency. In this study, those who were treated with estrogens differed from those who were not in age, age of menopause, duration of follow-up, height, weight, blood pressure, marital status, race, economic status, and gravidity, as well as in the frequency of the following diagnoses: atrophic vaginitis, bilateral oophorectomy, premature ovarian failure, hypopituitarism, gonadal dysgenesis, endocrine disease, hypertension, and osteoporosis [50].

A third study used a case–control design to investigate patients admitted to surgical services [52]. It compared cases with hip fractures to a control group of surgical patients, divided into those with and those without trauma. Cases were noted to be older, taller, and to have a lower body weight than the controls. The cases more frequently had undergone ovariectomy, breastfed fewer times and for fewer months, and were hypothyroid less frequently than the controls. When these factors were controlled for as confounding variables, the effect of estrogens was still apparent. However, as in the other studies, there was no information on how or why the decision was made to treat with or withhold estrogens.

A number of other nonexperimental studies published since then showed similar results [59,61,62,66–71]. Since then, the finding that estrogens have a beneficial effect on hip fractures has been confirmed in a large clinical trial, the Women's Health Initiative [72].

Anticoagulants for Prevention of Recurrent Venous Thromboembolism

The use of intravenous anticoagulants reduces the risk of recurrent venous thromboembolism [73], and the addition of oral anticoagulants to intravenous anticoagulants probably reduces the risk even further [74]. However, how long oral anticoagulant treatment should be continued had not been well studied. Most explicit advice from experts on the optimal duration of anticoagulation therapy was based on anecdotal experience [75,76]. Most of the data that were used to suggest the appropriate duration of therapy are derived from clinical observations in a single medical center [77–80]. They represent an accumulating case series. Over time, gradually patients' treatment has been prolonged. Thus, changes in the duration of treatment are intermingled with other changes in medical care over decades. In addition, the studies do not compare patients receiving treatments of different lengths but simply observe when most recurrences tend to occur. The investigators have assumed that treatment should be prolonged sufficiently to include that time when recurrences can be expected. Problems with these studies have been detailed [75,76].

As with the question of the effect of estrogens on bone fractures from osteoporosis, a randomized clinical trial would be the ideal design to address the question of the optimal duration of anticoagulation after venous thromboembolism, but such a study is difficult. After patients have been anticoagulated in the hospital and followed for a short time as outpatients, the risk of recurrence is sufficiently small that an enormous population would be needed to detect a difference in outcome due to differences in therapy. For years, the only randomized clinical trial in the literature that addressed this question compared six weeks of outpatient treatment to six months of treatment. No difference in recurrence rate between these two groups of patients was observed [81]. However, only 186 subjects were included, yielding a total of only seven recurrences. In addition, over half the study subjects had some known short-term risk factors for venous thromboembolism. These included pregnancy, use of oral contraceptives,

and recent surgery. Patients with these transient underlying risk factors might be expected to be less likely to benefit from longer-term anticoagulant therapy than patients with idiopathic disease.

The question of the optimal duration of anticoagulation was addressed in a retrospective cohort study performed using medical records review in the Northern California Kaiser Permanente Medical Program [82]. The study required the use of 10 years of data from this population of 1.6 million, or a total of 16 million patient-years of experience. A total of 3384 individuals were identified as being hospitalized for venous thromboembolism. Of these, 2473 suffered from idiopathic venous thromboembolism. Their clinical outcomes were evaluated according to how long they had been treated with oral anticoagulants. Using those treated with six weeks of therapy or less as a control group, prolongation of therapy beyond that point was found to increase the risk of major bleeding dramatically but to have no effect on recurrence rates. Unfortunately, very few of these episodes of venous thromboembolism were objectively confirmed, that is, they were clinical diagnoses only, as that was not the practice at Kaiser.

The feature of this study that allowed the investigators to overcome the problem of confounding by indication was that physician behavior regarding how long therapy was continued was essentially random (category 3.a.ii in Table 33.2). The choice of how long to treat became random, because there was no prior information on how long one should treat. In fact, the duration of treatment was relatively uniformly distributed across the years of follow-up, and the results were no different when one restricted the analysis to those who had their anticoagulation stopped because of hemorrhage, rather than at the option of their physician.

However, these results were not necessarily confirmed in subsequent randomized trials. A decade later, a multicenter trial in Sweden, with 897 patients with first episode of venous thromboembolism treated with oral anticoagulants and followed up for two years, found a significant difference in the incidence of recurrent venous thromboembolism between the six-week and six-month groups (18.1% vs 9.5%, respectively), and no significant difference in mortality or the incidence of major hemorrhage between the two treatment groups [83].

Several other recent studies also showed the benefit of longer duration of warfarin anticoagulant therapy. One randomized trial also showed that long-term low-dose warfarin therapy was effective in decreasing the subsequent risk of recurrence of idiopathic venous thromboembolism, in patients who had already received full-dose warfarin for a median of 6.5 months [84].

Lidocaine for Prevention of Death from Myocardial Infarction

In another study, the efficacy of lidocaine in preventing death from myocardial infarction was studied using a case–control design [85]. Among patients admitted to a coronary or intensive care unit for acute myocardial infarction, those who died were compared to an equal number of patients who survived. The controls were matched to the cases for age, gender, race, and date of hospitalization. Overall, lidocaine did not protect against death. Lidocaine was effective only when deaths attributable to ventricular arrhythmia were analyzed separately.

In this careful study, the investigators obviously were well aware of the risk of confounding by indication. They attempted to control for this by using the epidemiologic technique of stratification, that is, classifying patients according to their risk of dying from myocardial infarction, in order to control for this inequality of risk as a confounding variable. Thus, they treated the study as a category 3.b question in Table 33.2. Unfortunately, however, it is doubtful whether one can accurately and fully measure the basis for physicians' judgments about who they think is at high risk of death from myocardial

infarction. Similarly, it is unlikely that each individual's risk of dying from a myocardial infarction can be predicted, especially death by ventricular arrhythmia. Certainly, a classification according to just the presence or absence of congestive heart failure, as was used, is overly simplistic. In fact, the rates of death attributed to ventricular arrhythmia were virtually identical in those patients with and without congestive heart failure. Nevertheless, the results do coincide with those of a randomized clinical trial evaluating the efficacy of lidocaine in preventing primary ventricular fibrillation [86]. However, while the drug prevented the arrhythmia in that randomized clinical trial, it did not alter mortality.

Since then, there have been more than 20 randomized trials and four metaanalyses, indicating that lidocaine reduces ventricular fibrillation but, contrary to the results of the nonrandomized trials, increases mortality in acute myocardial infarction [87]. This was not confirmed in a subsequent paper, which reanalyzed the data from the 43 704 patients enrolled in GUSTO-I or GUSTO-IIb [88].

Anticoagulants for Prevention of Death from Myocardial Infarction

Whether anticoagulants can prevent death from myocardial infarction was addressed using randomized clinical trials [89]. However, the results were inconsistent and inconclusive, possibly because of problems of sample size. Thus, this question would appear to be a good candidate for a case–control study. Such a study was done [90], with the investigators treating this research question as if it were a category 3.b question in Table 33.2. However, as with the study of the effects of lidocaine on myocardial infarction, it is doubtful whether one can measure and quantitate precisely the risk of dying from a myocardial infarction at the time of the acute episode.

This study might have been more convincing if the investigators had identified the patients of practitioners who always used anticoagulants for their patients with myocardial infarctions, and then compared them to a control group of patients of practitioners who never used anticoagulants for their patients with myocardial infarctions. Inasmuch as the choice of therapy in these patients would not have been made on the basis of any perceived difference among the patients in their risk of dying from myocardial infarction, confounding by the indication would not be a problem. Of course, if the investigators had designed the study as we suggest, they then would have had to consider whether the physicians themselves were somehow a predictor of outcome, and whether this was consistently related to their philosophy of using anticoagulants, across multiple physicians. Thus, randomized trials are really needed to provide the answer to this question, and of course in recent years, with the advent of low molecular weight heparin and thrombolytic therapy, many have been forthcoming [91–96].

Generic versus Brand Name Drugs

Another potential use of nonexperimental designs to study the beneficial effects of drugs arose with the passage of the 1984 Waxman–Hatch Act in the US. Generic drugs can now be marketed after simple demonstration of bioequivalence, that is, equivalent bioavailability, in 18–24 normal adults [97]. However, it is not clear whether bioequivalence assures clinical equivalence, that is, equivalent efficacy and toxicity [98]. *Clinical inequivalence is more likely to be evident as a difference in beneficial effects than as a difference in adverse effects*. In developing a drug, dosages are sought which optimize drug efficacy. Toxicity, other than idiosyncratic or allergic reactions, usually occurs at higher doses and concentrations than needed for efficacy. Modest variations in the plasma concentration of the active drug, created by receiving the same dose in different preparations, are most likely, therefore, to be a problem for drug efficacy than for drug toxicity. Variations in plasma concentration are even more likely to be a problem for drug effectiveness and

cost-effectiveness. Even a simple change in the physical appearance of the drug could conceivably lead to a decrease in adherence and, thereby, effectiveness.

Studies designed to evaluate differences in efficacy among different preparations of the same drug require enormous sample sizes, as one would be searching for relatively small differences. However, such sample sizes can be achieved relatively easily and efficiently as part of nonexperimental pharmacoepidemiologic studies. Thus, the suggestion has been made that studies of clinical equivalence could possibly be carried out as postmarketing surveillance studies [98]. Confounding by the indication might be unlikely to be a problem because, as far as the physician is concerned, he or she is dealing with different products of the same drug, products that are theoretically interchangeable. The choice among the alternative therapies might be expected to be made irrespective of patient characteristics, but rather by the pharmacist on the basis of product availability – category 3.a.ii in Table 33.2.

A few pharmacoepidemiologic studies (unpublished) on the relative effectiveness of different preparations used for the same purpose were performed by Strom, using the COMPASS[R] database. These studies compared patients who were started on a brand name product and were switched to a generic product when it became available with patients who remained on the brand name product. The drugs studied were thioridazine, chlorpropamide, and slow-absorption theophylline. These studies naturally raise concerns about the ability to identify the actual product dispensed. Very few of the pharmacoepidemiologic approaches described in Part III of the book are able to identify the specific product dispensed. Often the approach does not even distinguish whether it is a brand name product or a generic product that is being used. Even when the distinction is made (for example, most Medicaid datasets use the National Drug Code to identify specifically the drug, the manufacturer, the dosage form, and the dose), one is inevitably left with questions about whether a brand name is being billed for while a generic drug is dispensed. In addition, such studies raise concerns about how to define the clinical outcome variable. For example, how is drug efficacy reflected in a claims database? The studies described above used proxy outcomes such as number of physician visits, number of hospitalizations, and use of adjunctive therapy to obtain an estimate of drug efficacy.

Using these outcomes, the investigators first analyzed the baseline data, comparing the experience, prior to switching, of those who ultimately switched to generic products to the experience of those who did not switch. In each of the three studies, the future switchers were different from the future nonswitchers, prior to the switch. Thus, it appears that patients who were to be switched to generic products were different from patients who stayed on the brand name products: confounding by indication was indeed operating. Because of this, no analyses of efficacy after the switch were performed. Parenthetically, because of this, and questions about the uncertain interpretability of the clinical outcomes, it was elected not to publish the results of these papers.

Cost-effectiveness Studies

An important category of studies of beneficial drug effects includes studies of their cost-effectiveness. These measure the resources necessary to achieve a particular beneficial outcome, and thus have two main study variables – one that is clinical and one that is economic [99–102]. For example, one could perform a cohort study comparing treated patients to untreated patients, and determine whether the clinical outcomes they experience and the cost of the medical care they subsequently receive is different. In such a study, one would need to consider the possibility of confounding by the indication for both the clinical outcome and the cost

variables. It should be noted that the indication may have different effects on the clinical outcomes and the costs. Thus, while performing the clinical outcome assessment, one needs to consider and, potentially, quantify the implications of the indication for the treatment on the clinical outcome variable. In contrast, while performing the cost assessment, one needs to consider and, potentially, quantify the cost implications of the indication on both the clinical outcomes and the costs. The subject of health economics as applied to drug use is discussed in more detail in Chapter 34.

Vaccines

Nonexperimental study designs have been widely used to evaluate the efficacy of vaccines. Specifically, case–control studies have been used to explore the efficacy of pneumococcal vaccine [83,84,103–106], rubella vaccine [107,108], measles vaccine [109–113], *Haemophilus influenzae* type b polysaccharide vaccine [114–125], oral poliovirus vaccine [126,127], meningococcus vaccine [128–130], Japanese encephalitis vaccine [131,132], BCG vaccine in protecting against tuberculosis [133–140], diphtheria toxoid vaccine [141], mumps vaccine [142], and leprosy [143,144]. Cohort studies have been used to explore the efficacy of *Haemophilus influenzae* type b polysaccharide vaccine [116], measles vaccine [117,145], and pertussis vaccine [146,147].

Again, studies like these should ideally be conducted as randomized clinical trials. However, the relative infrequency of the diseases that the above vaccines are designed to prevent, particularly in populations which are partly vaccinated, make use of this design difficult, although not impossible. In fact, in one situation, a new Japanese encephalitis vaccine manufactured in China was studied for efficacy using a case–control design [131], while a study of its safety, conducted by the same authors, used a randomized clinical trial design [132]. In considering the applicability of nonexperimental study designs, the relatively indiscriminate use of such vaccines places the study in category 2 of Table 33.2. Patients who receive these vaccines differ from those who do not in their socioeconomic status, access to medical care, and physicians' attitudes towards vaccines. However, for most vaccines, an individual physician is not likely to give only some of his/her eligible patients the vaccine, withholding it from other eligible patients. Thus, patients receiving vaccines are not likely to differ from those who do not get the vaccine, at least in their physicians' perceptions about the patients' risk of contracting these diseases. Nonexperimental studies of such questions should produce valid results, therefore. Indeed, as is evident from the large number of examples, this is becoming a standard and accepted approach. We refer the interested reader to some methodologic papers on the subtleties of designing nonexperimental studies of vaccine efficacy [148–154].

Other Examples

Other analogous work using case–control study designs has explored the effectiveness of bicycle safety helmets in preventing face injuries [155,156], antibiotic prophylaxis in preventing postdental infective endocarditis [157,158], beta-blockers in preventing mortality in patients with acute myocardial infarction [159], beta-blockers and incident coronary artery events [160], etc.

The Future

Clinicians have long recognized the value of clinical observations and nonexperimental research. Much of our current knowledge about the usefulness of medical interventions is based on information that is nonexperimental. However, the information that observational techniques generate cannot be accepted uncritically. Perhaps in reaction to the limitations of nonexperimental studies, some scientists have insisted that "the

randomized clinical trial (RCT) is the only scientifically reliable method for assessment of the efficacy (and risks) of most clinical treatments" [27]. Sackett *et al*. argued "… to keep up with the clinical literature … discard at once all articles on therapy that are not randomized trials" [161]. In light of the analysis presented above, this position seems too simplistic and far reaching. If overbearing, it results in clinically necessary and potentially available information being uncollected and unused. The proper balance in attitude about the value of these approaches probably lies somewhere between the two extremes. To quote Sir Austin Bradford Hill, one of the developers of the randomized trial: "Any belief that the controlled trial is the only way (to study therapeutic efficacy) would mean not only that the pendulum had swung too far but that it had come right off its hook [162]."

Many investigators are now applying nonexperimental designs to studies of beneficial drug effects, including comparing active treatments with each other (see Chapter 26). With the passage of the 21st Century Cures Act, the US Congress embraced the use of "real-life evidence" and the fact that it might provide data on drug effects, including benefit. However, careful attention needs to be paid to the possibility of confounding by the indication. Some approaches to this problem are now available (see Chapter 43), and hopefully more will be available in the future. However, when confounding by indication can be addressed, clinical observations and nonexperimental research can be used. The results of nonexperimental research are unlikely to be as powerful or as convincing as those of experimental research. We are not suggesting that nonexperimental studies be used as replacements for experimental studies. However, when an experimental study is deemed to be unnecessary, unethical, infeasible, or too costly relative to the expected knowledge to be gained, there might be reasonable alternatives.

References

1 Wallerstein RO, Condit PK, Kasper CK, Brown JW, Morrison FR. Statewide study of chloramphenicol therapy and fatal aplastic anemia. *JAMA* 1969; **208**: 2045–50.

2 Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first dose phenomenon. *BMJ* 1976; **2**: 1293–4.

3 Boston Collaborative Drug Surveillance Group. Regular aspirin intake and acute myocardial infarction. *BMJ* 1974; **1**: 440–3.

4 Kimmel SE, Berlin JA, Reilly M, *et al*. The effects of nonselective non-aspirin nonsteroidal anti-inflammatory medications on the risk of nonfatal myocardial infarction and their interaction with aspirin. *J Am Coll Cardiol* 2004; **43**: 985–90.

5 Strom BL, Melmon KL, Miettinen OS. Postmarketing studies of drug efficacy: Why? *Am J Med* 1985; **78**: 475–80.

6 Anonymous. Ibuprofen (Motrin) – a new drug for arthritis. *Med Lett Drugs Ther* 1974; **16**: 109–10.

7 Wilson TW, Falk KJ, Labelle JL, Nguyen KB. Effect of dosage regimen on natriuretic response to furosemide. *Clin Pharmacol Ther* 1975; **18**: 165–9.

8 Anonymous. Clonidine (Catapres) for hypertension. *Med Lett Drugs Ther* 1975; **17**: 45–6.

9 Shen SW, Bressler R. Clinical pharmacology of oral antidiabetic agents. *N Engl J Med* 1977; **296**: 787–93.

10 Webb-Johnson DC, Andrews JL. Bronchodilator therapy. *N Engl J Med* 1977; **297**: 758–64.

11 Hollister LE. Tricyclic antidepressants. *N Engl J Med* 1978; **299**: 1106–9.

12 Syriopoulou V, Scheifele D, Howie V, Ploussard J, Sloyer J, Smith AL. Incidence of

ampicillin-resistant Hemophilus influenzae in otitis media. *J Pediatr* 1976; **89**: 839–41.

13 Schwartz R, Rodriquez W, Khan W, Ross S. The increasing incidence of ampicillin-resistant Hemophilus influenzae. *JAMA* 1978; **239**: 320–3.

14 Vestal RE. Drug use in the elderly: a review of problems and special considerations. *Drugs* 1978; **16**: 358–82.

15 Morselli PL, Principi N, Tognoni G, *et al*. Diazepam elimination in premature and full term infants, and children. *J Perinat Med* 1973; **1**: 133–41.

16 Chan RA, Bennes EJ, Hoeprich PD. Gentamicin therapy in renal failure: a nomogram for dosage. *Ann Intern Med* 1972; **76**: 773–8.

17 Hansten PD. *Drug Interactions*, 4th edn. Philadelphia: Lea & Febiger, 1979.

18 *Physicians' Desk Reference*, 33rd edn. Oradell: Medical Economics, 1979.

19 Parijs J, Joossens JV, van der Linden L, Verstreken G, Amery AK. Moderate sodium restriction and diuretics in the treatment of hypertension. *Am Heart J* 1973; **85**: 22–34.

20 Doering PL, Stewart RB. The extent and character of drug consumption during pregnancy. *JAMA* 1978; **239**: 843–6.

21 Boethius G. Recording of drug prescriptions in the county of Jämtland, Sweden. II. Drug exposure of pregnant women in relation to course and outcome of pregnancy. *Eur J Clin Pharmacol* 1977; **12**: 37–43.

22 Degenhardt KH, Kerken H, Knörr K. Drug usage and fetal development: preliminary evaluations of a prospective investigation. *Adv Exp Med Biol* 1972; **27**: 467–79.

23 Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? *Clin Pharmacol Ther* 1983; **34**: 1–7.

24 Psaty BM, Koepsell TD, Siscovick D, *et al*. An approach to several problems in using large databases for population-based case-control studies of the therapeutic efficacy and safety of anti-hypertensive medicines. *Stat Med* 1991; **10**: 653–62.

25 Carson JL, Strom BL, Soper KA, West SL, Morse ML. The association of nonsteroidal anti-inflammatory drugs with upper gastrointestinal tract bleeding. *Arch Intern Med* 1987; **147**: 85–8.

26 Strom BL, Carson JL, Morse ML, West SL, Soper KA. The effect of indication on hypersensitivity reactions associated with zomepirac sodium and other nonsteroidal antinflammatory drugs. *Arthritis Rheum* 1987; **30**: 1142–8.

27 Juhl E, Christensen E, Tygstrup N. The epidemiology of the gastrointestinal randomized clinical trial. *N Engl J Med* 1977; **296**: 20–2.

28 Rosenbaum PR. *Observational Studies*. Springer Series in Statistics. New York: Springer-Verlag, 1995, pp. 200–24.

29 D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; **17**: 2265–81.

30 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.

31 Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003;**158**: 280–7.

32 Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**: 703–8.

33 Strom BL. Are estrogens effective in preventing fractures from post-menopausal osteoporosis? In: Melmon KL, ed. *Drug Therapeutics: Concepts For Physicians*. New York: Elsevier-North Holland, 1982, pp. 67–80.

34 Lagrelius A. Treatment with oral estrone sulphate in the female climacteric. III. Effects on bone density and on certain biochemical parameters. *Acta Obstet Gynecol Scand* 1981; **60**: 481–8.

35 Lindsay R, Hart DM, Aitken JM, MacDonald EB, Anderson JB, Clarke AC. Long-term prevention of postmenopausal osteoporosis by oestrogen. Evidence for an increased bone mass after delayed onset of oestrogen treatment. *Lancet* 1976; **1**: 1038–41.

36 Lindsay R, Hart DM, MacLean A, Clark AC, Kraszewski A. Bone response to termination of oestrogen treatment. *Lancet* 1978; **1**: 1325–7.

37 Ziel HK, Finkle WD. Increased risk of endometrial cancer among users of conjugated estrogens. *N Engl J Med* 1975; **293**: 1167–70.

38 Antunes CM, Strolley PD, Rosenshein NB, *et al*. Endometrial cancer and estrogen use. Report of a large case-control study. *N Engl J Med* 1979; **300**: 9–13.

39 Hutchinson TA, Polansky SM, Feinstein AR. Post-menopausal oestrogens protect against fractures of hip and distal radius. *Lancet* 1979; **2**: 705–9.

40 Weiss NS, Ure CL, Ballard JH, Williams AR, Daling JR. Decreased risk of fractures of the hip and lower forearm with postmenopausal use of estrogen. *N Engl J Med* 1980; **303**: 1195–8.

41 Johnson RE, Specht EE. The risk of hip fracture in postmenopausal females with and without estrogen drug exposure. *Am J Public Health* 1981; **71**: 138–44.

42 Wallach S, Henneman PH. Prolonged estrogen therapy in post-menopausal women. *JAMA* 1959; **171**: 1637–42.

43 Gordon GS, Picchi J, Roof BS. Antifracture efficacy of long-term estrogens for osteoporosis. *Trans Assoc Am Phys* 1973; **86**: 326–32.

44 Burch JC, Byrd BF, Vaughn WK. The effects of long-term estrogen on hysterectomized women. *Am J Obstet Gynecol* 1974; **118**: 778–82.

45 Burch JC, Byrd BF, Vaughn WK. The effects of long-term estrogen administration to women following hysterectomy. *Front Horm Res* 1975; **3**: 208–14.

46 Nachtigall LE, Nachtigall RH, Nachtigall RD, Beckman EM. Estrogen replacement therapy I. A 10-year prospective study in the relationship to osteoporosis. *Obstet Gynecol* 1979; **53**: 277–81.

47 Nachtigall LE, Nachtigall RH, Nachtigall RD, Beckman EM. Estrogen replacement therapy II. A prospective study in the relationship to carcinoma and cardiovascular and metabolic problems. *Obstet Gynecol* 1979; **54**: 74–9.

48 Nordin BE, Horsman A, Crilly RG, Marshall DH, Simpson M. Treatment of spinal osteoporosis in postmenopausal women. *BMJ* 1980; **280**: 451–5.

49 Paganini-Hill A, Ross RK, Gerkins VR, Henderson BE, Arthur M, Mack TM. Menopausal estrogen therapy and hip fractures. *Ann Intern Med* 1981; **95**: 28–31.

50 Hammond CB, Jelovsek FR, Lee KL, Creasman WT, Parker RT. Effects of long-term estrogen replacement therapy I. *Metabolic effects. Am J Obstet Gynecol* 1979; **133**: 525–36.

51 Riggs BL, Seeman E, Hodgson SF, Taves DR, O'Fallon WM. Effect of the fluoride/calcium regimen on vertebral fracture occurrence in postmenopausal osteoporosis. Comparison with conventional therapy. *N Engl J Med* 1982; **306**: 446–50.

52 Krieger N, Kelsey JL, Holford TR, O'Connor T. An epidemiologic study of hip fracture in postmenopausal women. *Am J Epidemiol* 1982; **116**: 141–8.

53 Hooyman JR, Melton LJ, Nelson AM, O'Fallon WM, Riggs BL. Fractures after rheumatoid arthritis. *Arthritis Rheum* 1984; **27**: 1353–61.

54 Ettinger B, Genant HK, Cann CE. Long-term estrogen replacement therapy prevents bone loss and fractures. *Ann Intern Med* 1985; **102**: 319–24.

55 Wasnich RD, Ross PD, Heilbrun LK, Vogel JM, Yano K, Benfante RJ. Differential effects of thiazide and estrogen upon bone mineral content and fracture prevalence. *Obstet Gynecol* 1986; **67**: 457–62.

56 Michaelsson K, Baron JA, Farahmand BY, *et al*. Hormone replacement therapy and risk of hip fracture: population based case-control study. The Swedish Hip Fracture Study Group. *BMJ* 1998; **316**: 1858–63.

57 Michaëlsson K, Baron JA, Johnell O, Persson I, Ljunghall S. Variation in the efficacy of hormone replacement therapy in the prevention of hip fracture. Swedish Hip Fracture Study Group. *Osteoporos Int* 1998; **8**: 540–6.

58 Vestergaard P, Rejnmark L, Mosekilde L. Fracture reducing potential of hormone replacement therapy on a population level. *Maturitas* 2006; **54**: 285–93.

59 Strom BL, Schinnar R, Weber AL, *et al*. Case-control study of postmenopausal hormone replacement therapy and endometrial cancer. *Am J Epidemiol* 2006; **164**: 775–86.

60 Rebbeck TR, Troxel AB, Wang Y, *et al*. Estrogen sulfation genes, hormone replacement therapy, and endometrial cancer risk. *J Natl Cancer Inst* 2006; **98**: 1311–20.

61 Lacey JV Jr, Leitzmann MF, Chang SC, *et al*. Endometrial cancer and menopausal hormone therapy in the National Institutes of Health-AARP Diet and Health Study cohort. *Cancer* 2007; **109**: 1303–11.

62 Epstein E, Lindqvist PG, Olsson H. A population-based cohort study on the use of hormone treatment and endometrial cancer in southern Sweden. *Int J Cancer* 2009; **125**: 421–5.

63 Crandall C. Low-dose estrogen therapy for menopausal women: a review of efficacy and safety. *J Womens Health* 2003; **12**: 723–47.

64 Phillips LS, Langer RD. Postmenopausal hormone therapy: critical reappraisal and a unified hypothesis. *Fertil Steril* 2005; **83**: 558–66.

65 Sammartino A, Cirillo D, Mandato VD, di Carlo C, Nappi C. Osteoporosis and cardiovascular disease: benefit-risk of hormone replacement therapy. *J Endocrinol Invest* 2005; **28**(10 Suppl): 80–4.

66 Nelson HD, Rizzo J, Harris E, *et al*. Osteoporosis and fractures in postmenopausal women using estrogen. *Arch Intern Med* 2002; **162**: 2278–84.

67 Randell KM, Honkanen RJ, Kroger H, Saarikoski S. Does hormone-replacement therapy prevent fractures in early postmenopausal women? *J Bone Miner Res* 2002; **17**: 528–33.

68 Cauley JA, Zmuda JM, Ensrud KE, Bauer DC, Ettinger B. Timing of estrogen replacement therapy for optimal osteoporosis prevention. *J Clin Endocrinol Metab* 2002; **86**: 5700–5.

69 Mosekilde L, Beck-Nielsen H, Sorensen OH, *et al*. Hormonal replacement therapy reduces forearm fracture incidence in recent postmenopausal women – results of the Danish Osteoporosis Prevention Study. *Maturitas* 2000; **36**: 181–93.

70 Honkanen RJ, Honkanen K, Kroger H, Alhava E, Tuppurainen M, Saarikoski S. Risk factors for perimenopausal distal forearm fracture. *Osteoporos Int* 2000; **11**: 265–70.

71 Grodstein F, Stampfer MJ, Falkeborn M, Naessen T, Persson I. Postmenopausal hormone therapy and risk of cardiovascular disease and hip fracture in a cohort of Swedish women. *Epidemiology* 1999; **10**: 476–80.

72 Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; **288**: 321–33.

73 Barritt DW, Jordan SC. Anticoagulant drugs in the treatment of pulmonary embolism. *Lancet* 1960; **1**: 1309–12.

74 Coon WW, Willis PW. Recurrence of venous thromboembolism. *Surgery* 1973; **73**: 823–7.

75 Acheson L, Speizer FE, Tager I. Venous thromboembolism duration of anticoagulant therapy (letter). *N Engl J Med* 1975; **293**: 879.

76 Salzman EW. Venous thrombosis. duration of anticoagulant therapy (letter). *N Engl J Med* 1975; **293**: 879.

77 Coon WW, Mackenzie JW, Hodgson PE. A critical evaluation of anticoagulant therapy in peripheral venous thrombosis and pulmonary embolism. *Surg Gynecol Obstet* 1958; **106**: 129–36.

78 Coon WW, Willis PW, Symons MJ. Assessment of anticoagulant treatment of venous thromboembolism. *Ann Surg* 1969; **170**: 559–68.

79 Coon WW, Willis PW. Thromboembolic complications during anticoagulant therapy. *Arch Surg* 1972; **105**: 209–12.

80 Coon WW, Willis PW. Recurrence of venous thromboembolism. *Surgery* 1973; **73**: 823.

81 O'Sullivan EF. Duration of anticoagulant therapy in venous thrombo-embolism. *Med J Aust* 1972; **2**: 1104–7.

82 Petitti DB, Strom BL, Melmon KL. Duration of warfarin anticoagulant therapy and the probabilities of recurrent thromboembolism and hemorrhage. *Am J Med* 1986; **81**: 255–9.

83 Schulman S, Rhedin AS, Lindmarker P, *et al*. A comparison of six weeks with six months of oral anticoagulant therapy after a first episode of venous thromboembolism. Duration of Anticoagulation Trial Study Group. *N Engl J Med* 1995; **332**: 1661–5.

84 Ridker PM, Goldhaber SZ, Danielson E, *et al*. Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. *N Engl J Med* 2003; **348**: 1425–34.

85 Horwitz RI, Feinstein AR. Improved observational method for studying therapeutic efficacy. Suggestive evidence that lidocaine prophylaxis prevents death in acute myocardial infarction. *JAMA* 1981; **246**: 2455–9.

86 Lie KI, Wellens HJ, van Capelle FJ, Durrer D. Lidocaine in the prevention of primary ventricular fibrillation. A double-blind, randomized study of 212 consecutive patients. *N Engl J Med* 1974; **291**: 1324–6.

87 Sadowski ZP, Alexander JH, Skrabucha B, *et al*. Multicenter randomized trial and a systematic overview of lidocaine in acute myocardial infarction. *Am Heart J* 1999; **137**: 792–8.

88 Alexander JH, Granger CB, Sadowski Z, *et al*. Prophylactic lidocaine use in acute myocardial infarction: incidence and outcomes from two international trials. The GUSTO-I and GUSTO-IIb Investigators. *Am Heart J* 1999; **137**: 799–805.

89 Hirsh J. Effectiveness of anticoagulants. *Semin Thromb Hemost* 1986; **12**: 21–37.

90 Horwitz RI, Feinstein AR. The application of therapeutic-trial principles to improve the design of epidemiologic research: a case-control study suggesting that anticoagulants reduce mortality in patients with myocardial infarction. *J Chronic Dis* 1981; **34**: 575–83.

91 Collins R, MacMahon S, Flather M, *et al*. Clinical effects of anticoagulant therapy in suspected acute myocardial infarction: systematic overview of randomised trials. *BMJ* 1996; **313**: 652–9.

92 Yusuf S, Wittes J, Friedman L. Overview of results of randomized clinical trials in heart disease I. Treatments following myocardial infarction. *JAMA* 1988; **260**: 2088–93.

93 Chalmers TC, Matta RJ, Smith H Jr, Kunzler AM. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977; **297**: 1091–6.

94 Peto R. Clinical trial methodology. *Biomed Pharmacother* 1978; **28**(special issue): 24–36.

95 Mitchell JRA. Anticoagulants in coronary heart disease: retrospect and prospect. *Lancet* 1981; **2**: 1014–17.

96 Timmis GC, Mammen EF, Ramos RG, *et al*. Hemorrhage vs rethrombosis after thrombolysis for acute myocardial infarction. *Arch Intern Med* 1986; **146**: 667–72.

97 Mattison N. Pharmaceutical innovation and generic drug competition in the USA: effects of the Drug Price Competition and Patent Term Restoration Act of 1984. *Pharmaceut Med* 1986; **1**: 177–85.

98 Strom BL. Generic drug substitution – revisited. *N Engl J Med* 1987; **316**: 1456–62.

99 Mohle-Boetani JC, Schuchat A, Plikaytis BD, Smith JD, Broome CV. Comparison of prevention strategies for neonatal group B streptococcal infection. A population-based economic analysis. *JAMA* 1993; **270**: 1442–8.

100 Pearle MS, Roehrborn CG. Antimicrobial prophylaxis prior to shock wave lithotripsy in patients with sterile urine before treatment: a meta-analysis and cost-effectiveness analysis. *Urology* 1997; **49**: 679–86.

101 Chaix C, Durand-Zaleski I, Alberti C, Brun-Buisson C. Control of endemic methicillin-resistant *Staphylococcus aureus*: a cost-benefit analysis in an intensive care unit. *JAMA* 1999; **282**: 1745–51.

102 Eymann R, Chehab S, Strowitzki M, Steudel WI, Kiefer M. Clinical and economic consequences of antibiotic-impregnated cerebrospinal fluid shunt catheters. *J Neurosurg Pediatr* 2008; **1**: 444–50.

103 Shapiro ED, Clemens JD. A controlled evaluation of the protective efficacy of pneumococcal vaccine for patients at high risk of serious pneumococcal infections. *Ann Intern Med* 1984; **101**: 325–30.

104 Mills OF, Rhoads GG. The contribution of the case-control approach to vaccine evaluation: Pneumococcal and Haemophilus influenzae type B PRP vaccines. *J Clin Epidemiol* 1996; **49**: 631–6.

105 Vila-Corcoles A, Salsench E, Rodriguez-Blanco T, *et al*. Clinical effectiveness of 23-valent pneumococcal polysaccharide vaccine against pneumonia in middle-aged and older adults: a matched case-control study. *Vaccine* 2009; **27**: 1504–10.

106 Barricarte A, Castilla J, Gil-Setas A, *et al*. Effectiveness of the 7-valent pneumococcal conjugate vaccine: a population-based case-control study. *Clin Infect Dis* 2007; **44**: 1436–41.

107 Greaves WL, Orenstein WA, Hinman AR, Nersesian WS. Clinical efficacy of rubella vaccine. *Pediatr Infect Dis* 1983; **2**: 284–6.

108 Strassburg MA, Greenland S, Stephenson TG, *et al*. Clinical effectiveness of rubella vaccine in a college population. *Vaccine* 1985; **3**: 109–12.

109 Killewo J, Makwaya C, Munubhi E, Mpembeni R. The protective effect of measles vaccine under rountine vaccination conditions in Dar Es Salaam, Tanzania: a case-control study. *Int J Epidemiol* 1991; **20**: 508–14.

110 Mahomva AI, Moyo IM, Mbengeranwa LO. Evaluation of a measles vaccine efficacy during a measles outbreak in Mbare, City of Harare Zimbabwe. *Cent Afr J Med* 1997; **43**: 254–6.

111 Mudzamiri WS, Peterson DE, Marufu T, Biellik RJ, L'Herminez M. Measles vaccine efficacy in Masvingo district, Zimbabwe. *Cent Afr J Med* 1996; **42**: 195–7.

112 Kabir Z, Long J, Reddaiah VP, Kevany J, Kapoor SK. Non-specific effect of measles vaccination on overall child mortality in an area of rural India with high vaccination coverage: a population-based case-control study. *Bull World Health Organ* 2003; **81**: 244–50.

113 Akramuzzaman SM, Cutts FT, Hossain MJ, *et al*. Measles vaccine effectiveness and risk factors for measles in Dhaka, Bangladesh. *Bull World Health Organ* 2002; **80**: 776–82.

114 Wenger JD, Pierce R, Deaver KA, Plikaytis BD, Facklam RR, Broome CV and the Haemophilus Influenzae Vaccine Efficacy Study Group. Efficacy of Haemophilus influenzae type b polysaccharide-diphtheria toxoid conjugate vaccine in US children aged 18–59 months. *Lancet* 1991; **338**: 395–8.

115 Greenberg DP, Vadheim CM, Bordenave N, *et al*. Protective efficacy of Haemophilus influenzae type b polysaccharide and conjugate vaccines in children 18 months of age and older. *JAMA* 1991; **265**: 987–92.

116 Kaiser Permanente Pediatric Vaccine Study Group. Efficacy of Haemophilus influenzae type b capsular polysaccharide vaccine. *Pediatr Infect Dis* 1988; **7**: 149–56.

117 Sutcliffe PA, Rea E. Outbreak of measles in a highly vaccinated secondary school population. *Can Med Assoc J* 1996 **155**: 1407–13.

118 Harrison LH, Broome CV, Hightower AW, *et al*. A day care-based study of the efficacy of Haemophilus b polysaccharide vaccine. *JAMA* 1988; **260**: 1413–18.

119 Shapiro ED, Murphy TV, Wald ER, Brady CA. The protective efficacy of Haemophilus b polysaccharide vaccine. *JAMA* 1988; **260**: 1419–22.

120 Osterholm MT, Rambeck JH, White KE, *et al*. Lack of efficacy of Haemophilus b polysaccharide vaccine in Minnesota. *JAMA* 1988; **260**: 1423–8.

121 Black SB, Shinefield HR, Hiatt RA, Fireman BH, and the Kaiser Permanente Pediatric Vaccine Study Group. Efficacy of Haemophilus influenzae type b capsular polysaccharide vaccine. *Pediatr Infect Dis J* 1988; **7**: 149–56.

122 Bower C, Condon R, Payne J, Burton P, Watson C, Wild B. Measuring the impact of conjugate vaccines on invasive Haemophilus influenzae type b infection in Western Australia. *Aust N Z J Public Health* 1998; **22**: 67–72.

123 Muhlemann K, Alexander ER, Weiss NS, Pepe M, Schopfer K. Risk factors for invasive Haemophilus influenzae disease among children 2–16 years of age in the vaccine era, Switzerland 1991–1993. The Swiss H. Influenzae Study Group. *Int J Epidemiol* 1996; **25**: 1280–5.

124 McVernon J, Andrews N, Slack MP, Ramsay ME. Risk of vaccine failure after Haemophilus influenzae type b (Hib) combination vaccine with acellular pertussis. *Lancet* 2003; **361**: 1521–3.

125 McVernon J, Johnson PD, Pollard AJ, Slack MP, Moxon ER. Immunologic memory in Haemophilus influenzae type b conjugate vaccine failure. *Arch Dis Child* 2003; **88**: 379–83.

126 Balraj V, John TJ, Thomas M, Mukundan S. Efficacy of oral poliovirus vaccine in rural communities of North District, India. *Int J Epidemiol* 1990; **19**: 711–14.

127 Deming MS, Jaiteh KO, Otten MW Jr, *et al*. Epidemic poliomyelitis in The Gambia following the control of poliomyelitis as an endemic disease. II. Clinical efficacy of trivalent oral polio vaccine. *Am J Epidemiol* 1992; **135**: 393–408.

128 Rheingold AL, Broome CV, Hightower AW, *et al*. Age-specific differences in duration of clinical protection after vaccination with meningococcal polysaccharide A vaccine. *Lancet* 1985; **2**: 114–21.

129 Rosenstein N, Levine O, Taylor JP, *et al*. Efficacy of meningococcal vaccine and barriers to vaccination. *JAMA* 1998; **279**: 435–9.

130 Bose A, Coen P, Tully J, Viner R, Booy R. Effectiveness of meningococcal C conjugate vaccine in teenagers in England. *Lancet* 2003; **361**: 675–6.

131 Hennessy S, Lui Z, Tsai TF, *et al*. Effectiveness of live-attenuated Japanese encephalitis vaccine (SA14-14-2): a case-control study. *Lancet* 1996; **347**: 1583–6.

132 Liu ZL, Hennessy S, Strom BL, *et al*. Short-term safety of live-attenuated Japanese encephalitis vaccine (SA14-14-2): results of a 26,239-subject randomized clinical trial. *J Infect Dis* 1997; **176**: 1366–9.

133 Miceli I, Colaiacovo D, Kantor IN, Peluffo G. Evaluation of the effectiveness of BCG vaccination using the case-control method. *Dev Biol Stand* 1986; **58**: 293–6.

134 Shapiro C, Cook N, Evans D, *et al*. A case-control study of BCG and childhood tuberculosis in Cali, Colombia. *Int J Epidemiol* 1985; **14**: 441–6.

135 Young TK, Hershfield ES. A case-control study to evaluate the effectiveness of mass BCG vaccination among Canadian Indians. *Am J Public Health* 1986; **76**: 783–6.

136 Houston S, Fanning A, Soskolne CL, Fraser N. The effectiveness of bacillus calmette-guerin

(BCG) vaccination against tuberculosis. *Am J Epidemiol* 1990; **131**: 340–8.

137 Filho VW, de Castilho EA, Rodrigues LC, Huttly SRA. Effectiveness of BCG vaccination against tuberculous meningitis: a case-control study in Sao Paulo, Brazil. *Bull WHO* 1990; **68**: 69–74.

138 Patel A, Schofield F, Siskind V, Abrahams E, Parker J. Case-control evaluation of a school-age BCG vaccination programme in subtropical Australia. *Bull WHO* 1991; **69**: 425–33.

139 Sirinavin S, Chotpitayasunondh T, Suwanjutha S, Sunakorn P, Chantarojanasiri T. Protective efficacy of neonatal Bacillus Calmette-Guérin vaccination against tuberculosis. *Pediatr Infect Dis J* 1991; **10**: 359–65.

140 Sharma RS, Srivastava DK, Singh AA, *et al*. Epidemiological evaluation of BCG vaccine efficacy in Delhi – 1989. *J Commun Dis* 1989; **21**: 200–6.

141 Bisgard KM, Rhodes P, Hardy IR, *et al*. Diphtheria toxoid vaccine effectiveness: a case-control study in Russia. *J Infect Dis* 2000; **181**(Suppl 1): 84–7.

142 Fu C, Liang J, Wang M. Matched case-control study of effectiveness of live, attenuated S79 mumps virus vaccine against clinical mumps. *Clin Vaccine Immunol* 2008; **15**: 1425–8.

143 Fine PEM, Ponninghaus JM, Maine N, Clarkson JA, Bliss L. Protective efficacy of BCG against leprosy in Northern Malawi. *Lancet* 1986; **2**: 499–502.

144 Muliyil J, Nelson KE, Diamond EL. Effect of BCG on the risk of leprosy in an endemic area: a case control study. *Int J Lepr* 1991; **59**: 229–36.

145 Ramsay ME, Moffatt D, O'Connor M. Measles vaccine: a 27-year follow-up. *Epidemiol Infect* 1994; **112**: 409–12.

146 Isomura S. Clinical studies on efficacy and safety of an acellular pertussis vaccine in Aichi prefecture, Japan. *Dev Biol Stand* 1991; **73**: 37–42.

147 Report from the PHLS Epidemiological Research Laboratory and 21 area health authorities. Efficacy of pertussis vaccination in England. *BMJ* 1982; **285**: 357–9.

148 Smith PG. Retrospective assessment of the effectiveness of BCG vaccination against tuberculosis using the case-control method. *Tubercle* 1982; **63**: 23–35.

149 Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *Int J Epidemiol* 1984; **13**: 87–93.

150 Clemens JD, Shapiro ED. Resolving the pneumococcal vaccine controversy: are there alternatives to clinical trials? *Rev Inf Dis* 1984; **6**: 589–600.

151 Orenstein WA, Bernier RH, Dondero TJ, *et al*. Field evaluation of vaccine efficacy. *Bull WHO* 1985; **63**: 1055–68.

152 Smith PG. Evaluating interventions against tropical diseases. *Int J Epidemiol* 1987; **16**: 159–66.

153 Farrington CP. Quantifying misclassification bias in cohort studies of vaccine efficacy. *Stat Med* 1990; **9**: 1327–37.

154 Joseph KS. A potential bias in the estimation of vaccine efficacy through observational study designs (letter). *Int J Epidemiol* 1989; **18**: 729–31.

155 Thompson DC, Thompson RS, Rivara FP, Wolf ME. A case-control study of the effectiveness of bicycle safety helmets in preventing facial injury. *Am J Public Health* 1990; **80**: 1471–4.

156 Cummings P, Rivara FP, Thompson DC, Thompson RS. Misconceptions regarding case-control studies of bicycle helmets and head injury. *Accid Anal Prev* 2006; **38**: 636–43.

157 Imperiale TF, Horwitz RI. Does prophylaxis prevent postdental infective endocarditis? A controlled evaluation of protective efficacy. *Am J Med* 1990; **88**: 131–6.

158 Strom BL, Abrutyn E, Berlin JA, *et al*. Dental and cardiac risk factors for infective

endocarditis. A population-based, case-control study. *Ann Intern Med* 1998; **129**: 761–9.

159 Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**: 630–8.

160 Psaty BM, Koepsell TD, LoGerfo JP, Wagner EH, Inui TS. Beta-blockers and primary prevention of coronary heart disease in patients with high blood pressure. *JAMA* 1989; **261**: 2087–94.

161 Sackett DL, Haynes RB, Tugwell PT. *Clinical Epidemiology. A Basic Science for Clinical Medicine*. Boston: Little, Brown, 1985.

162 Hill AB. Reflections on the controlled trial. *Ann Rheum Dis* 1966; **25**: 107–13.

34

Pharmacoeconomics: The Economics of Pharmaceuticals

Kevin A. Schulman

Clinical Excellence Research Center, Department of Medicine, Stanford University, Stanford, CA, USA

Modern medicine is defined by the use of pharmaceutical products. The treatment of many conditions would be unthinkable without these essential products. Yet, the price of pharmaceutical products and their economic impact on the healthcare system remain controversial. In this chapter, we explore the economics of pharmaceutical products.

Clinical Problems to Be Solved by Pharmacoepidemiologic Research

A patient walks into a pharmacy with a prescription written by their physician for a new medication. The patient has health insurance from their employer that includes a specific prescription drug benefit. The pharmacist dispenses the medication and asks the patient to pay for their portion of the cost. The pharmacy then bills the insurer for the balance of the cost. These types of transactions are typical, and they occur daily at every pharmacy in the United States (and similar transactions occur in other countries). Yet behind this simple occurrence is the complex economics of healthcare and the pharmaceutical market. This chapter reviews the economics of the pharmaceutical market and provides some insights into potential ways in which this market may evolve in the future.

The Economics of Pharmaceutical Products

The science of drug development and assessment has been well described in the literature and throughout this book. Drug development proceeds through stages of discovery, optimization, clinical development, regulatory review, and launch with postmarket assessment. The economics of this process brings together concepts of finance, health economics, and behavioral economics in a manner that is truly unique.

Drug development starts with an investment in science. Historically, public grant funds through the National Institutes of Health or the National Science Foundation, or private funds through programs like the Howard Hughes Medical Institute, would support fundamental science that might be years or even generations away from translation into medical products. Academic researchers and the pharmaceutical industry would use the insights from this work to begin an effort at translation, moving from

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

fundamental science to specific interventions for specific diseases. This work could still be publicly funded, but might be more likely to be funded through private resources such as pharmaceutical firms, or even applied science efforts such as the Bill and Melinda Gates Foundation. These efforts serve to translate biology into drug targets, identifying potential pathways to alter the target through the identification of small molecules or biologics that have (hopefully) unique effects on the target. The discovery of a candidate drug would lead to the filing of a patent, an opportunity for the inventor to own the rights to the discovery and to preclude others from practicing the invention (under a provision called the Bayh–Dole Act, universities own the patent rights to discoveries even if the work was funded using federal grants).

The patent rights are critical in the next stage of drug development. This is when the process moves from the laboratory to human testing. This is the most expensive step in drug development, and for the most part the work is privately funded. Investors justify their investment in a molecule with the opportunity for financial returns resulting from their ownership of the molecule through the patent. This transition from public to private support is challenging for many discoveries. The "Valley of Death" is deemed the gap between science that is funded by public grants and the ability to attract private investment to the development of a molecule.

Clinical testing of pharmaceuticals is carefully regulated by the Food and Drug Administration (FDA) and other global regulatory bodies. For most products, regulatory authorities require proof of safety and efficacy of products before they are approved for sale. Clinical testing can require up to a decade to complete and can require more than $1 billion in direct outlays [1].

At the end of drug development, with product approval by the FDA, the manufacturer can set a price and market the product. Prices set by manufacturers reflect their significant investment in clinical development, and the inherent risk they were required to assume, but also considerations of market access or barriers to full reimbursement for patients. The price can reflect the marginal cost of producing a product, but often this is a relatively minor consideration. The prices of specialty pharmaceutical products can be extraordinary, reaching $475 000 per patient for Novartis's CAR-T therapy [2]. Prices also vary across markets, with a 30-day supply of Janssen's Xarelto® priced at $48 in South Africa, $102 in Switzerland, and $292 in the United States, or 400 mg of Genentech's Avastin® priced at $956 in South Africa, $1752 in Switzerland, and $3930 in the United States [3]. Not only are drug prices high, but cancer therapies have experienced significant price growth. In one analysis, the monthly cost of oncology products has increased from approximately $100 as recently as 1980 to $10 000 by 2010 [4].

A lot of attention has been focused on the impact of health insurance on the prices of pharmaceutical products. Historically, prescription drugs were relatively affordable, and so were paid for by patients. As medications became more effective, the concept of prescription drug insurance began to develop. In 1960 in the United States, 96% of prescription drug spending was out of pocket by individuals. By 1980, out-of-pocket spending was still 71% of total spending. By 1990, it was down to 57%, 2000 to 28%, 2010 to 18%, and 2015 to 14% [5]. Prescription drug coverage has led to a transformation of the pharmaceutical market. Over this same period, the prescription drug market has grown from $2.7 billion in 1960 to $462 billion in gross sales in 2016 (resulting in net pharmaceutical sales to manufacturers of $318 billion) [6].

Insurance is a mechanism for sharing risk across individuals. Generally, insurance works best when the occurrence being insured is infrequent, can be catastrophic to the individual, and is not influenced by the individual or organization being insured. In insurance markets characterized by these conditions, insurance can be a

relatively inexpensive proposition. Health insurance has different characteristics. We use healthcare services frequently, and with medications even more frequently. While some healthcare costs can be catastrophic, not all are. While paying for a monthly medication is not enjoyable, for most people it is not financially catastrophic. Finally, consumption of healthcare services is inherently influenced not only by individuals (do you want to go to the clinician for that bad cold or sprained ankle?), but also by pharmaceutical manufacturers and healthcare systems commercializing medicinal drugs and services. Consequently, health insurance has become a very expensive insurance product. In truth, most health insurance combines the idea of prepayment for usual healthcare services with a catastrophic medical benefit (to some extent, high-deductible health plans try to separate out these two different elements of healthcare financing).

The idea of making a risky financial investment in order to garner a financial return is not unique to the pharmaceutical industry. The oil and gas industry faces a similar economic proposition in oil exploration, with risky new leases requiring years of development and the outlay of billions of dollars before realizing any return. One fundamental difference between pharmaceuticals and the petroleum industry is that the market sets the price for the products in the oil industry, not the oil producer. Also, consumers pay directly for the products in the oil industry while they often have insurance to help pay for pharmaceutical products. The final difference with the petroleum industry is that consumers may make very different purchase decisions in healthcare settings than they make in other aspects of their life.

In the next section, we explore the impact of health insurance on pricing decisions by manufacturers. We then explore an emerging area of economics, behavioral economics, and its ability to help us understand the unique price-inelastic behavior of segments of the healthcare market.

Health Economics

Moral Hazard

Health economists have long been worried about the economic impact of health insurance on the patterns of consumption of healthcare due to a concept called "moral hazard." Moral hazard describes the change in individual behavior between conditions of self-pay and conditions of third-party payment. Kenneth Arrow was awarded the Nobel Prize in economics for developing this framework [7], and Mark Pauly further developed the theory to focus on demand [8].

The basic framework is easy to understand. We all make purchases based on our concept of value. We generally make purchases of goods or products for $1.00 when we perceive that they offer $1.00 worth of value. This concept of value is an individual determination: we all have our own tastes, preferences, and needs which form our assessment of value.

Third-party payment alters this fundamental calculus. Consider going out to dinner with a group of friends. After the menu is passed around, you notice items of lower and higher price, say salad and steak. You can approach payment in one of two ways: individual checks or splitting the check. If you all decide on individual checks before you order, you may decide to purchase the lower-cost salad since you are on a budget. However, what happens if you decide to split the check before you order? You may be worried that everyone else at the table is likely to order the higher-priced steak, and you will have to pay your share of their higher-priced meals. Since you are paying for their steak, why not order your own steak so at least you get the benefit of the higher price you will pay for dinner? In this simple illustration, your behavior changes between self-payment and third-party payment models.

Health insurance is one form of third-party payment. Under health insurance, rather than

Table 34.1 Perception of value to the patient.

	Product 1	Product 2	Product 3	Product 4	Product 5	Product 6	Product 7
Value	1.50	1.25	1.0	.75	.50	.25	.10
Cost (No Insurance)	$1.00	$1.00	$1.00	$1.00	$1.00	$1.00	$1.00
Cost (Insurance)	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20

Note: Cost (no insurance) assumes only cash payments for the product. Cost (insurance) assumes the product is covered by an insurance policy with a 20% co-insurance requirement.

paying the full cost of medical products, you pay only a co-payment (fixed amount), or co-insurance (a percentage payment) for medical products. As illustrated in Table 34.1, products 1 through 3 offer at least $1.00 of value for $1.00 of cost. In a self-payment model, you would be expected to purchase only products 1 through 3 since only these products have a value of $1.00. In an insurance model, however, you only pay the co-payment of $0.20. Now, products 1 through 6 offer value equal to or greater than the $0.20 copayment, so using the same rule (only buying products that offer value greater than or equal to the price you pay), you would purchase products 1 through 6. Again, behavior changes under conditions of third-party payment. While many economists have argued that health insurance increases the overall cost of healthcare due to these changes in demand [9], there is also the concept of good moral hazard where people can purchase goods or products through insurance that would otherwise be unaffordable [10]. It is possible to develop a direct estimate of the increase in prescription drug costs between those with and without insurance [11].

To this point, the discussion has focused on the impact of moral hazard on the demand for healthcare products. However, the impact of moral hazard also extends to the supply side of healthcare [12–14]. While much of the literature examines the impact of moral hazard on the provision of services, there is also an impact on the price of products. Given insurance, the suppliers of high-value products can realize that products are perceived as significantly underpriced since insured patients only consider the out-of-pocket costs. Applying a value framework to pricing can lead manufacturers to raise their prices to meet the value threshold rather than simply developing a price to meet their internal financial expectations. This supply-side moral hazard effect on the price of pharmaceutical products has been much less discussed in the literature [15–17].

Again, going back to the basic example of product 1 in Table 34.1, this product provides great value to patients under conditions of self-payment and even more under conditions of third-party payment. Sophisticated suppliers will notice these conditions. In a competitive market, suppliers will have little ability to influence the welfare surplus enjoyed by patients in this example since the price is determined by the market and is driven by entry and exit of firms. However, there are circumstances when suppliers have power to influence prices, especially in healthcare. Suppliers can have market power when they have a barrier to market entry such as a patent awarded to a pharmaceutical manufacturer or a product developed for a niche category, such as an orphan drug, which is too small to attract competition. In these cases, suppliers can increase the price of product 1 based on value. If they decide to price at the total value of the product, they could raise

the price from $1.00 to $1.50 to capture the full value to patients. Under our conceptual model, this pricing strategy would be attractive to patients even in a cash pay market. However, under conditions of third-party payment, suppliers can consider an even more aggressive pricing strategy by considering that patients measure value against their co-insurance, not the full cost of the product. Under these conditions, suppliers can raise the price to $7.50 while consumers would have a cost-share of $1.50, or an amount equal to the value they expect to receive from the therapy. As a result of supply-side moral hazard, the cost increased from $1.00 to $7.50 in this simple example. Co-payment coupons or patient assistance programs can exacerbate this effect by artificially decreasing the amount individuals have to pay. This "discount" on out-of-pocket payments can allow suppliers even more latitude to raise prices under this framework.

The supply-side implications of moral hazard are potentially significant. Beyond the short-term impacts on patients, this effect can have longer-term effects by distorting the drug development portfolio. In Table 34.2, we imagine a manufacturer with a simple two-product portfolio, with each product having equal development costs and market price. In analyzing their options, the firm invests in the opportunity with the largest market size [18].

However, in Table 34.3, under conditions of market power, manufacturers can consider the question of value of the therapy to patients in setting a price. In this case, they chose to undertake development of product B despite its smaller market size.

Thus, the supply-side effects of moral hazard can be seen in both the prices of products in the marketplace and the portfolio of drug products available on the market.

Table 34.2 Pharmaceutical market under conditions of supply-side moral hazard.

Product	Cost of development	Size of target market	Price	Revenue
A	$50 000 000	20 000	$10 000	$200 M
B	$50 000 000	10 000	$10 000	$100 M

Note: Cost of development – out-of-pocket dollar costs of development (assumption). Size of target market – number of accessible candidates for therapy considering incidence and prevalence of underlying condition. Price – market price for the product (net price to manufacturer). Revenue – net revenue from the product (price times market size).

Behavioral Economics

This concept of patients being risk averse is consistent with the idea of buying health insurance in the first place. Buying health insurance is seen as a risk-averse financial decision. People pay some money annually for health insurance to avoid the potential financial consequences

Table 34.3 Pharmaceutical market under conditions of market power.

Product	Cost of development	Size of target market	Value of therapy	Price	Revenue
A	$50 000 000	20 000	1	$10 000	$200 M
B	$50 000 000	10 000	5	$50 000	$500 M

Note: Cost of development – out-of-pocket dollar costs of development (assumption). Size of target market – number of accessible candidates for therapy considering incidence and prevalence of underlying condition. Value of therapy – perception of value to the patient (in dollar equivalents). Price – value price for the product (net price to manufacturer). Revenue – net revenue from the product (price times market size).

associated with the rare risk of becoming severely ill. Consumers may even buy certain policies with limits on things that are not important to them when they are healthy – narrow networks of providers, for example, or limits on the drug formulary for specialty pharmaceutical products. However, buying health insurance is not the same as buying healthcare. Whether the risk-averse decision-making approach to buying insurance carries over to making treatment decisions for healthcare products or services is an open question.

Let's consider a clinical scenario. Assume an otherwise healthy patient comes into a physician's office. They feel great, have a full social and work life, exercise regularly, and have a lot to look forward to. Given a history of smoking in the past, the physician had ordered a chest X-ray. Unfortunately, the chest X-ray shows that the patient has a spot on their lung. After further work-up, it is found to be lung cancer that has spread. This otherwise healthy person now has a life-threatening condition. Obviously, this is a significant loss in life expectancy for the patient. How do they react to the shock of their diagnosis? They seek treatment for their condition. In this case, the patient will accept a treatment which has any chance of restoring their health, irrespective of the side effects of the therapy. They definitely don't ask about the cost of treatment. Under conditions of loss, the way that patients make decisions changes from how they felt about future potential treatment choices when they bought their health insurance policy [1].

This idea that people make different decisions under conditions of gains and losses earned Daniel Kahneman the Nobel Prize in Economics in 2002 "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" [19]. (He collaborated with Amos Tversky in developing prospect theory, but Amos passed away before the prize was awarded.) Under conditions of gains, we are risk averse, and under conditions of loss we are risk seeking. When a 70-year-old patient refuses a flu shot because of her concerns that she may get sick from the shot, she is under a condition of gain (full health) and is being risk averse. The unfortunate patient with lung cancer is an example of decision making under conditions of loss. The application of this framework to treatment choices by patients with life-threatening diseases helps to explain the apparently risk-seeking behavior of patients [20,21]. This study of the psychology of decision making in real-world settings has been called behavioral economics.

More recently, Kahneman and others have focused on the role of emotion in decision making [22]. They have developed a framework which considers two different ways of making decisions, System 1 and System 2. System 1 decision processes are autonomous decision-making efforts that represent our "gut" or emotional response to an uncertain situation. System 1 processes easily incorporate societal attitudes and are subject to many systematic flaws [23]. System 2 decision processes are more data driven and analytical, but have a high cognitive burden. In the normal course of events, we make most decisions using the System 1 framework, despite its limitations, so that we minimize our cognitive burden in making simple choices or completing simple job tasks. However, we have System 2 processes available for more complex decision making. Importantly, in a heightened emotional state, we generally rely on System 1 processes for decision making. This can be critically important in understanding medical decision making, where patients (or their loved ones) can experience significant anxiety arising from the care process or the diagnosis itself, or can be in a heightened emotional state from the experience of the symptoms of the illness,

especially when suffering from a disease with an acute presentation.

While the role of loss can make patients appear to be risk seeking in making treatment choices, we have suggested that the role of emotion can also lead to the same type of decision making by patients [24].

Public perceptions and communication around a diagnosis such as cancer can have significant emotional overtones. Patients "fight a cancer diagnosis" and society is in the midst of a "war" on cancer [25]. In light of this, heart disease was responsible for almost 300 000 deaths in women in 2013 [26] while only 40 000 women died of breast cancer that same year [27]. This is not to say that breast cancer is not a terrible disease, but why do we spend so much time talking about an illness that has one-tenth the mortality of another, more prevalent disease?

It is interesting to think that this was not always the case. In his fascinating book *The Emperor of All Maladies*, Siddhartha Mukkherjee profiles the career of Dr Sidney Farber [28]. Farber was very interested in trying to develop treatments for childhood leukemia which was almost uniformly fatal. However, there was little funding for cancer research at the time. He started the Jimmy Fund in 1948, highlighting the plight of his young patients to raise money for cancer research [29]. The success of this strategy of creating cancer as an emotional disease to attract research support continues to this day. Most recently, the Cancer Moonshot concept was used to develop support for the Precision Medicine Initiative at NIH [30].

While these efforts have been successful in raising support for cancer research (and genomic medicine), they shape the perceptions of patients with one of these diseases. Receiving a diagnosis of a disease with a higher emotional tone may lead to patients finding it difficult to engage in System 2 thinking when making treatment choices. Both emotion and loss can help explain treatment choices for patients with life-threatening conditions. The result is that there is demand for therapies where the risks can seemingly outweigh the benefits of therapy, or for therapies where the benefits seem especially modest given the high cost. These frameworks can help to explain the decision-making process used by those such as cancer patients who have significant demand for therapies despite the tremendous cost and limited benefit of many agents.

An important aspect of this discussion is that under either of these decision-making frameworks, loss or emotion, patients are not actively required to consider cost in their decision-making process, nor might the cost of therapies negatively affect the decision-making processes of patients with life-threatening conditions. This suggests that under these conditions, demand for products could appear to be price inelastic. This lack of an impact of cost on decision making contrasts this type of decision making with moral hazard. It also suggests that mechanisms to address moral hazard (cost sharing, for example) would not be expected to influence demand for therapies by patients experiencing emotion or loss. In fact, knowing this, excessive cost sharing for these patients might even be considered unethical. The price of products for these patients would still be affected by the impact of moral hazard on supplier pricing decisions (price-inelastic demand might even exacerbate this effect).

At this juncture, it is too soon to know if the Precision Medicine Initiative will extend our emotional characterization of disease to conditions which were not previously considered a separate disease or extend conditions experienced in emotional terms such as cancer. It is unclear if the efforts to establish genetic causes of common conditions or to develop new labels for subsets of common diseases will increase the emotional tone associated with these conditions in the mind of the public [31].

Clinical Example: Treatment of Hepatitis C

The development of a new generation of pharmaceutical therapies for the treatment of chronic infection with hepatitis C was greeted as a significant scientific advance [32,33]. However, the payer community was in shock over the original price of $84 000 per patient for a 12-week course of therapy (or double that amount for 24 weeks) [34]. This set off a debate around the price of the therapy and led to significant challenges for many public and private payers [35]. With the high prevalence of chronic hepatitis C infection in the population, providing therapy for the entire population could have doubled health insurance premiums in the United States for a year [23].

Quickly, the debate around sofosbuvir turned to pharmacoeconomics as an attempt to understand the price. The pharmaceutical manufacturer, Gilead, argued that the high price was justified by the significant clinical benefit experienced by patients (although this argument was absent in a Senate investigation of internal Gilead discussions in setting the price for sofosbuvir [36]). Progression of hepatitis C infection led to the development of cirrhosis, or liver failure, and hepatocellular carcinoma, a fatal form of liver cancer, leading to requirements for hospitalization, liver transplant if a donor organ was available, or to death. Yet, the clinical data were early. There were no long-term outcomes studies showing effectiveness of treatment on reducing these complications of hepatitis C infection, just on sustained virologic response in short-term clinical trials.

In the United Kingdom, there is a requirement for a formal economic evaluation of the impact of a new therapy on healthcare costs and outcomes to better understand the value of therapy as part of the consideration of whether it should be included in the national formulary. The determination of value is achieved by assessing the incremental cost of the new therapy in comparison with the incremental clinical benefit resulting from treatment, as discussed in detail in the section on Methodologic Issues in the Pharmacoeconomic Assessment of Therapies. The result of this analysis is a cost-effectiveness ratio, or an assessment of the additional cost of achieving an additional quality-adjusted life-year (QALY) for patients. In the UK, £20 000 per QALY is considered a benchmark of good value, but this means that spending will increase since the drug requires an additional expenditure to receive the additional benefit.

A detailed evaluation by the UK National Institute for Health and Care Excellence (NICE) [37], considering an independent technology review and a price of £34 504 per standard course of therapy [38], concluded that for some indications sofosbuvir had a high likelihood of meeting a cost-effectiveness ratio of £20 000 per QALY (and an even higher likelihood of meeting a ratio of £30 000 per QALY). They recommended that NHS England try to find the resources required to implement this therapy.

In the United States, Gilead's first-year revenue for sofosbuvir was $10 billion, making it one of the top-grossing pharmaceutical products in the world [39]. However, the entry of additional novel therapies for hepatitis C led to price competition in the marketplace, with prices ranging from $26 400 to $62 500 per treatment course by 2017 [40]. Still, Gilead reported sales of hepatitis C products of $19.1 billion in 2015 and $14.8 billion in 2016 [41].

The pricing of sofosbuvir represents an interesting dilemma in modern health economics. Clearly, the company could not have set such a high price for their product in the absence of health insurance coverage. In fact, the high price actually helped patients receive better coverage as a catastrophic medical expense than they would have received under standard insurance coverage. The company, and physicians, perceive hepatitis C as a potentially life-threatening condition, leading to demand for the treatment.

How should patients, physicians, and payers make a choice about the value of this therapy?

Currently Available Solutions

Pharmacoeconomics

Pharmacoeconomic studies have been designed to meet the different information needs of healthcare purchasers and regulatory authorities [42–55]. Economic data from Phase III studies are used to support initial pricing of new therapies and are used in professional educational activities by pharmaceutical firms. Postmarketing economic studies are used to compare new therapies with existing therapies and increasingly to confirm the initial Phase III economic assessments of the product [56].

No single study can possibly provide all interested audiences with complete economic information about a new therapy. Thus, specific studies are undertaken to address economic concerns from specific perspectives, such as a postmarketing study of a new therapy from the perspective of a health maintenance organization (HMO). They may also be undertaken to assess the effect of therapy on specific cost categories, such as an assessment of the productivity costs of treatment, to provide data to federal governments in Europe, since these governments fund both the health insurance system and the disability system.

Across the globe, technology assessment agencies have been established to help provide or evaluate economic data as part of the reimbursement process [57–60]. In the UK, NICE provides guidance to the National Health Service [61]. In Germany, the Institute for Quality and Efficiency in Health Care (IQWiG) evaluates the effectiveness of drugs [62]. In the US, the Institute for Clinical and Economic Review (ICER) is a private organization publishing independent economic analyses of new pharmaceutical products [63].

Economic Evaluation and the Drug Development Process

The drug development process allows for timely collection of data that can be used to evaluate the costs and effects of pharmaceuticals early in their product life, with an opportunity for further data collection and evaluation once the product has been approved and marketed.

Clinical economics has been integrated throughout the development process, with goals that parallel the clinical development stages. Phase I and II studies are used to develop pilot economic data, such as estimates of the mean and variance estimates for costs, quality of life, and utilities for patients with a specific clinical syndrome. These studies are also used to perform pilot tests of data collection tools, including items in case report forms to prospectively capture resources used by patients who will be entered into the Phase III and postmarketing clinical trials. From these data, issues such as sample size and power for pharmacoeconomic studies can be assessed.

Incorporation of economic analyses as part of Phase III clinical trials is well established [64]. Phase III studies can include economic assessments of new therapies as a primary or secondary endpoint (i.e., an assessment of changes in the use of specific resource categories resulting from treatment, such as changes in length of hospital stay or hospitalization rates) [65–71].

Lastly, a wide variety of postmarketing economic studies can be performed. These include comparative effectiveness/efficiency trials (also known as “pragmatic” or “practical” trials) in which comparisons between products are made in more realistic settings with less restrictive protocols than those designed for Phase III safety and efficacy trials (see Chapter 32) [72]. These postmarketing studies may include assessments of the new therapy compared with “usual care” or with specific therapeutic agents. Again, the economic analysis can serve as a primary or secondary endpoint of the study.

Developing economic data as an endpoint in a clinical trial requires integrating pharmacoeconomics into the clinical development process. While there has been an increase in the number of trials that collect economic data, the challenge remains to ensure that pharmacoeconomic endpoints are considered sufficiently early in the clinical development process so that designing the economic protocol does not impede the process of designing the clinical trial. Economic analysis requires the establishment of a set of economic endpoints for study (e.g., direct costs, productivity, intangible costs to patients and caregivers, quality-of-life or preference measures for patients and caregivers), review of the clinical protocol to ensure that there are no economic biases in the design of the clinical trial – such as requirements for differential resource use between the treatment arms of the study – and development of the economic protocol. Ideally, the economic study will be integrated into the clinical protocol, and the economic data will be collected as part of a unified case report form for both clinical and economic variables.

Economic analysis faces further challenges depending on the indication for the product and the size of the clinical trial. For primary care products, such as hypertension or diabetes, clinical trials are global undertakings, raising questions about the generalizability of the economic results for individual payers. For specialty pharmaceutical products, the issue is the potential small size of the clinical trials affecting the ability to make a reasonable assessment of the economic impact of a therapy. For example, a novel exon-skipping treatment for Duchenne muscular dystrophy, eteplirsen, was an orphan drug tested in a dozen patients in a clinical trial with an endpoint of dystrophin production and performance on a six-minute walk test. There were no long-term data for clinical outcomes available from the trial [73].

In the following sections, we briefly review the research methods of pharmacoeconomics, discuss some methodologic issues that have confronted researchers investigating the economics of pharmaceuticals, and illustrate the usefulness of pharmacoeconomic research.

Methodologic Issues in the Pharmacoeconomic Assessment of Therapies

Techniques of Clinical Economics

Economists emphasize that costs are more than just transactions of currency. Cost represents the consumption of a resource that could otherwise be used for another purpose. The value of the resource is that of its next best use, which no longer is possible once the resource has been used. This value is called the resource's "opportunity cost." For example, the time it takes to read this chapter is a cost for the reader, because it is time that cannot be used again; the opportunity to use it for another purpose has been foregone. Good investments are made when the benefits of the investment (e.g., what you learn) are greater than or equal to the value of the opportunities you have foregone (e.g., what you would be doing if you were not reading this chapter).

In addition to the fact that not all costs involve a transaction of money, it is important to remember that, at least from the perspective of society as a whole, not all transactions of money should be considered costs. For example, monetary transactions that do not represent the consumption of resources (e.g., social security payments, disability payments, or other retirement benefits) are not costs by this definition. They simply transfer the right to consume the resources represented by the money from one individual to another.

In considering economic analysis of medical care, there are three dimensions of analysis, represented by the three axes of the cube in Figure 34.1 with which readers should become familiar. Along the X-axis are three types of economic analysis—cost identification,

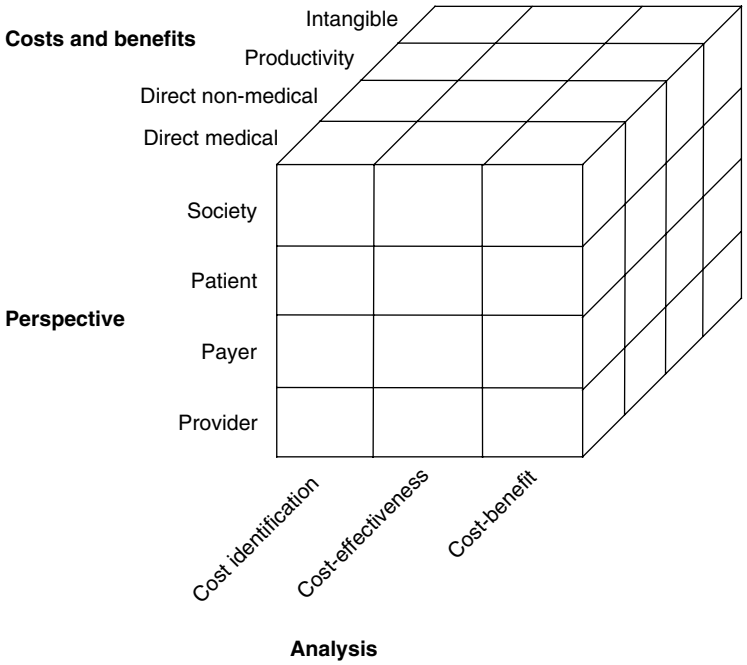


Figure 34.1 The three dimensions of economic evaluation of clinical care. *Source:* Bombardier C, Eisenberg J. Looking into the crystal ball: can we estimate the lifetime cost of rheumatoid arthritis? *J Rheumatol* 1985; **12**: 201–4. Reproduced with permission from the Journal of Rheumatology Publishing Company Ltd.

cost-effectiveness, and cost–benefit. Along the Y-axis are four points of view, or perspectives, that one may take in carrying out an analysis. One may take the point of view of society in assessing the costs and benefits of a new medical therapy. Alternatively, one may take the point of view of the patient, the payer, or the provider. Along the third axis, the Z-axis, are the types of costs and benefits that can be included in economic analysis of medical care. These costs and benefits, which will be defined below, include direct costs and benefits, productivity costs and benefits, and intangible costs and benefits.

Types of Analysis

Cost–Benefit Analysis

Cost–benefit analysis of medical care compares the cost of a medical intervention to its benefit. Both costs and benefits are measured in the same (usually monetary) units (e.g., dollars). These measurements are used to determine either the ratio of dollars spent to dollars saved or the net saving (if benefits are greater than costs) or net cost. All else being equal, an investment should be undertaken when its benefits exceed its costs.

The methods of cost–benefit analysis may be applied to evaluate the total costs and benefits of the interventions that are being compared by analyzing their cost–benefit ratios or their net benefits. Furthermore, the additional or "incremental" cost of an intervention (i.e., the difference in cost between a new therapy and conventional medical care) may be compared with its additional or "incremental" benefit. Incremental analysis is generally preferred to comparisons of totals because it allows the analyst to focus on the differences between any two treatment modalities. A zero-based budgeting approach would start from the base case. An

incremental analysis could overstate the value of a therapy if it compares a new therapy to an expensive but less effective therapy, while a zero-based budgeting approach would make the incremental costs and benefits of the new therapy more transparent.

One potential difficulty of cost–benefit analysis is that it requires researchers to express an intervention's costs and outcomes in the same units. Thus, monetary values must be associated with years of life lost and morbidity due to disease and with years of life gained and morbidity avoided due to intervention. Expressing costs in this way is difficult in healthcare analyses. Outcomes (treatment benefits) may be difficult to measure in units of currency. Such an exercise also raises methodologic and ethical questions of valuing human life differently across patients with different medical conditions, gender, occupations, or by age. Translating disease and treatment outcomes into monetary measures may be more difficult than translating them into clinical outcome measures, such as years of life saved or years of life saved adjusted for quality.

Cost-Effectiveness Analysis

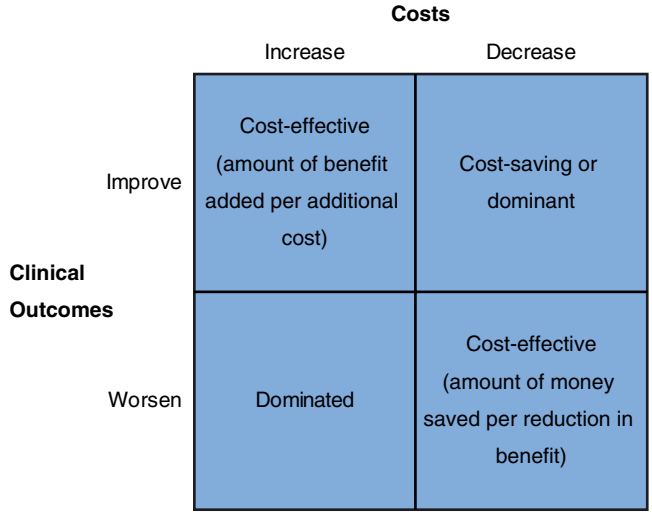
Cost-effectiveness analysis provides an alternative approach that avoids the dilemma of assessing the monetary value of health outcomes as part of the evaluation. While cost generally is still calculated only in monetary terms (e.g., dollars spent), effectiveness is determined independently and may be measured only in clinical terms, using any meaningful clinical unit. For example, one might measure clinical outcomes in terms of number of lives saved, complications prevented, or diseases cured. Alternatively, health outcomes can be reported in terms of a change in an intermediate clinical outcome, such as cost per percent change in blood cholesterol level. These results generally are reported as a ratio of costs to clinical benefits, with costs measured in monetary terms but with benefits measured in the units of the relevant outcome measure (for example, dollars per year of life saved).

When several outcomes result from a medical intervention (e.g., the prevention of both death and disability), cost-effectiveness analysis may consider these two outcomes together only if a common measure of outcome can be developed. Frequently, analysts combine different categories of clinical outcomes according to their desirability, assigning a weighted utility, or value, to the overall treatment outcome [3]. A utility weight is a measure of the patient's preferences for his/her health state or for the outcome of an intervention. The comparison of costs and utilities sometimes is referred to as cost–utility analysis, with the denominator expressed as QALYs.

In cost-effectiveness analysis, determination of value is based on the treatment's incremental costs and incremental effectiveness. In this approach, the analyst calculates the additional effect of one therapy compared with another (e.g., lives saved) per additional treatment dollar spent. Programs that cost less and demonstrate improved or equivalent treatment outcomes are said to be cost-saving or dominant and should always be adopted (Figure 34.2). Programs that cost more and are more effective are assessed by their cost-effectiveness ratio. Programs that cost more and have worse clinical outcomes are said to be dominated and should never be adopted. Programs that cost less and have reduced clinical outcomes may be adopted depending upon the magnitude of the changes in cost and outcome (the cost-effectiveness ratio assessing how much money would be saved in comparison with how little clinical outcomes were reduced).

As with the translation of clinical outcomes into monetary measures for cost–benefit analyses, there also are difficulties associated with combining different outcomes into a common measure in cost-effectiveness analysis. However, it is generally considered more difficult to translate all health benefits into monetary units for the purposes of cost–benefit analysis than to combine clinical outcomes measures. Thus,

Figure 34.2 Cost and clinical outcomes.



cost-effectiveness analysis is used more frequently than cost–benefit analysis in the medical care literature.

Net benefit, measured as a net monetary benefit or net health benefit, is a measure that combines estimates of incremental costs and incremental effectiveness (the components of an incremental cost-effectiveness ratio) with an estimate of the willingness-to-pay threshold. The willingness-to-pay threshold represents the maximum monetary outlay that would be acceptable for a one-unit gain in health benefit (e.g., $100 000 per QALY gained). Specifically, net monetary benefits are calculated by multiplying the willingness-to-pay threshold by the incremental effect (e.g., QALYs) and then subtracting the incremental cost. When net benefits are positive, the program should be adopted from a cost-effectiveness perspective. When net benefits are negative, the program is considered cost-inefficient and should not be adopted. An evaluation of net benefits differs from cost–benefit analysis because we do not directly assign monetary values to specific health outcomes, but instead use administratively determined valuations (e.g., $100 000 per QALY) [58,74,75]. Net benefit is particularly important for statistical evaluation of cost-effectiveness analysis (including sample size calculation and direct testing of economic value by use of patient-level data).

Cost Identification Analysis

An even less complex approach than cost–benefit or cost-effectiveness analysis would be simply to enumerate the costs involved in medical care and to ignore the outcomes that result from that care. This approach is known as cost identification analysis, by which the researcher can determine alternative ways of providing a service. The analysis might be expressed in terms of the cost per unit of service provided. For example, a cost identification study might measure the cost of a course of antibiotic treatment, but it would not calculate the clinical outcomes (cost-effectiveness analysis) or the value of the outcomes in units of currency (cost–benefit analysis). Cost identification studies, which include comparisons among different treatments based upon their costs alone, are appropriate only if treatment outcomes or benefits are equivalent among the therapies being evaluated.

Sensitivity Analysis

Most cost–benefit and cost-effectiveness studies require large amounts of data that may vary in reliability and validity, and could affect the overall results of the study. This is especially the case when models are developed for the economic analysis using secondary data sources, when data collection is performed retrospectively, or when critical data elements are unmeasured or unknown. Sensitivity analysis is a set of procedures in which the results of a study are recalculated using alternate values for some of the study's variables in order to test the sensitivity of the conclusions to these altered specifications. Such an analysis can yield several important results by demonstrating the independence or dependence of a result on particular assumptions, establishing the minimum or maximum values of a variable that would be required to affect a recommendation to adopt or reject a program, and identifying clinical or economic uncertainties that require additional research. In general, sensitivity analyses are performed on variables that have a significant effect on the study's conclusions but for which values are uncertain.

Types of Costs

Another dimension of economic analysis of clinical practice illustrated by Figure 34.1 is the evaluation of costs of a therapy. Economists consider three types of costs: direct, productivity, and intangible.

Direct Medical Costs

The direct medical costs of care usually are associated with monetary transactions and represent costs incurred during the provision of care. Examples of direct medical costs include payments for purchasing a pharmaceutical product, payments for physicians' fees, salaries of allied health professionals, or purchases of diagnostic tests. Because the charge for medical care may not accurately reflect the resources consumed, accounting or statistical techniques may be needed to determine direct costs [50,76–80].

Direct Nonmedical Costs

Monetary transactions undertaken as a result of illness or healthcare to detect, prevent, or treat disease are not limited to direct medical costs. There is another type of cost that is often overlooked: direct nonmedical costs. These costs are incurred because of illness or the need to seek medical care. They include the cost of transportation to the hospital or physician's office, the cost of special clothing needed because of the illness, the cost of hotel stays for receiving medical treatment at a distant medical facility, and the cost of special housing (e.g., modification of a home to accommodate an ill individual). Direct nonmedical costs, which are generally paid out of pocket by patients and their families, are just as much direct costs as are expenses that are more usually covered by third-party insurance plans.

Productivity Costs

In contrast to direct costs, productivity costs, sometimes referred to as indirect costs, do not stem from transactions for goods or services. Instead, they represent the cost of morbidity (e.g., time lost from work) or mortality (e.g., premature death leading to removal from the workforce). They are costs because they represent the loss of opportunities to use a valuable resource, a life, in alternative ways. A variety of techniques are used to estimate productivity costs of illness or healthcare [81–85]. Sometimes, as with varicella vaccination [86], the productivity costs of an illness are substantially greater than the direct costs of the illness [87,88].

Intangible Costs

Intangible costs are those of pain, suffering, and grief. These costs result from medical illness itself and from the services used to treat

the illness. They are difficult to measure as part of a pharmacoeconomic study, though they are clearly considered by clinicians and patients in considering potential alternative treatments. Although investigators are developing ways to measure intangible costs – such as willingness-to-pay analysis whereby patients are asked to place monetary values on intangible costs [46] – at present these costs are often omitted in clinical economics research.

Perspective of Analysis

The third axis in Figure 34.1 is the perspective of an economic analysis of medical care. Costs and benefits can be calculated with respect to society's, the patient's, the payer's, and the provider's points of view. A study's perspective determines how costs and benefits are measured, and the economist's strict definition of costs (the consumption of a resource that could otherwise be used for another purpose) no longer may be appropriate when perspectives different from that of society as a whole are used. For example, a hospital's cost of providing a service may be less than its charge. From the hospital's perspective, then, the charge could be an overstatement of the resources consumed for some services. However, if the patient has to pay the full charge, it is an accurate reflection of the cost of the service to the patient. Alternatively, if the hospital decreases its costs by discharging patients early, the hospital's costs may decrease but patients' costs may increase because of the need for increased outpatient expenses that are not covered by their health insurance plan.

Because costs will differ depending on the perspective, the economic impact of an intervention will be different from different perspectives. To make comparisons of the economic impact across different interventions, it is important for all economic analyses to adopt a similar perspective. It has been recommended that, as a base case, all analyses adopt a societal perspective [89]. The cost to society is the opportunity cost, the value of the opportunities foregone because the resource has consumed. Society's perspective usually is taken by measuring the consumption of real resources, including the loss of potentially productive human lives. As already noted, this cost does not count transfer payments, such as social security benefits. (From the point of view of the Social Security Administration, however, these payments would be a cost, because the perspective of the Social Security Administration is not the perspective of society.) If an intervention is not good value for money from the societal perspective, it would not be a worthwhile intervention for society, even if the intervention has economic advantages for other stakeholders.

Nevertheless, conducting the economic analysis from other perspectives, in addition to the societal perspective, is important. This is because the costs of medical care may not be borne solely by the same parties who stand to benefit from it. Economic analysis of medical care often raises vexing ethical problems related to equity, distribution of resources, and responsibility for the health of society's members [90,91]. Economic analyses from multiple perspectives shed light on the equity issues associated with new interventions.

In summary, economic analysis of medical technology or medical care evaluates a medical service by comparing its dollar cost with its dollar benefit (cost–benefit), by measuring its dollar cost in relation to its outcomes (cost-effectiveness), or simply by tabulating the costs involved (cost identification). Direct costs are generated as services are provided. In addition, productivity costs should be considered, especially in determining the benefit of a service that decreases morbidity or mortality. Finally, the perspective of the study determines the costs and benefits that will be quantified in the analysis, and sensitivity analyses test the effects of changes in variable specifications for estimated measures on the results of the study.

Currently Available Solutions: Imatinib Case Study

The previous sections of this chapter dealt with the principles of clinical economics and methodologic issues surrounding the economic analysis of pharmaceutical products. This section presents a case study that illustrates the practical application of these methods to the evaluation of pharmaceuticals. The following case demonstrates an approach to the economic evaluation of a new cancer therapy, imatinib (marketed as Gleevec® by Novartis).

Chronic myeloid leukemia (CML) is a malignant disorder of hematopoietic stem cells reported to account for 15–20% of adult leukemia cases. The disease presents with a relatively asymptomatic chronic phase but can be followed by accelerated and blast phases. While the disease has a median life expectancy of eight years, the life expectancy in advanced phases drops to months without treatment.

Chronic myeloid leukemia is an interesting disease. From a biological perspective, it results from a translocation between chromosomes 9 and 22, leading to the "Philadelphia chromosome" that is found in most of these patients. This translocation leads to the formation of a new gene, BCR-ABL, which produces a unique fusion protein, a tyrosine kinase. This protein includes effects such as activation of mitogenic signaling and inhibition of apoptosis leading to uncontrolled growth of white blood cells [92].

A unique collaboration between industry and academia searched for a kinase inhibitor that would be effective at targeting the BCR-ABL product. This eventually led to the discovery of the molecule that became imatinib [93].

Economic evaluation was an early consideration with this therapy. From the sponsor's perspective, CML was a limited market of approximately 5000 patients in the US [94]. However, if the drug was as effective as was hoped, the clinical benefit might help support internal expectations for a high target price for the therapy.

The International Randomized Interferon versus STI571 Study (IRIS) was an open-label trial that compared the efficacy of imatinib (the molecule was originally called signal transduction inhibitor [STI] 571) versus interferon (IFN-alpha) plus low-dose cytarabine (IFN+LDAC) in 1106 patients from 16 countries who were newly diagnosed with chronic-phase CML [95]. INF+LDAC had been the standard of care before imatinib was discovered but the regimen was not well tolerated and had limited clinical effectiveness. Thus, oncologists were excited about the potential for a novel therapy. However, given that the trial was open-label, the study suffered from significant patient cross-over. The study had a median follow-up of 19 months. While imatinib was generally well tolerated, 12.3% of patients in the imatinib group discontinued study medication while receiving first-line therapy (compared with 31.6% of patients in the IFN+LDAC group). However, the drug looked to perform as anticipated. At 18 months, the rate of complete cytogenetic response (CCyR) to first-line therapy was estimated at 76.2% for imatinib, compared with 14.5% for IFN+LDAC. Estimated rates of progression to accelerated phase or blast crisis were 3.3% in the imatinib group and 8.5% in the IFN+LDAC group. Overall, 2.0% of the patients randomized to receive imatinib and 57.5% randomized to receive IFN+LDAC crossed over to the alternate treatment after failing first-line therapy. This pattern of cross-over between study arms actually diminished the ability to detect an effect of imatinib in intention-to-treat analyses (see Chapter 32).

The economic investigators sought to estimate the incremental cost-effectiveness of imatinib compared with IFN+LDAC as the first-line treatment for patients with newly diagnosed chronic-phase CML [67].

Study Methods

The economic analysis of the IRIS study required construction of a model to assess the costs and benefits of therapies in both study arms. Data sources for this economic evaluation included data collected in IRIS and supplemental data from the literature. Given the significant cross-over in the clinical study, this usually challenging activity was even more complicated.

The economic analysis had to consider the use of imatinib as first-line therapy. The researchers therefore set up two treatment pathways. For patients on imatinib, they assumed that patients receiving imatinib as first-line therapy could switch to IFN+LDAC and then to hydroxyurea until disease progression. For the comparison group, patients receiving IFN+LDAC as first-line therapy, they assumed that the patients started therapy with INF+LDAC and then switched to hydroxyurea on discontinuation of IFN+LDAC until progression.

Since disease generally progressed from chronic phase to an advanced phase, the researchers developed separate estimates of resource use and quality of life for patients in each phase of their illness. For patients in the chronic phase, they developed different estimates of resource use and quality of life for each treatment arm, but assumed that both groups received the same therapies and had the same outcomes upon disease progression.

Since this was a model, best practice is to develop a framework for a broad sensitivity analysis to test critical assumptions in the model. In this case, the analysis explicitly incorporated the uncertainty associated with each parameter and provided the capability to test assumptions regarding efficacy, survival, duration of treatment, resource use, costs, quality-of-life weights, and discount rates. The analysis was conducted from the perspective of the healthcare system and considered only direct medical costs. In the base-case analysis, cost and survival estimates were discounted at 3% per year; discounting is an approach to understanding that we generally express more value for wealth or health today than at some point in the future. Thus, benefits today are perceived to be "worth more" than benefits in the future. To make current and future values equivalent, we can discount future values to show how nominal "future" values relate to "present" values today [2].

Survival Estimates

Given the high rates of treatment cross-over in the IRIS study, estimation of treatment benefit was a significant challenge. Survival for the two years after the initiation of treatment for chronic-phase CML was based on data from IRIS for patients receiving imatinib as first-line therapy. However, given the high rates of cross-over, the two-year survival rate for the IFN+LDAC arm could not be estimated from the IRIS trial. Instead, a historical control was used – a randomized trial conducted by the Italian Cooperative Study Group on CML for patients receiving IFN+LDAC [96].

While these analyses provided estimates of survival for the first two years of therapy, there was little observed progression during this time period since the chronic phase of CML can last up to eight years. However, there was an important biomarker of treatment effectiveness given the unique biology of CML. A complete cytogenetic response (CCyR) was reported to occur when there were no Philadelphia chromosomes found in white blood cells on a bone marrow biopsy. This was an important biomarker of treatment effect since the chromosome abnormality was felt to be the underlying cause of CML by producing the BCR-ABL protein.

Fortunately, there were survival data on a cohort of 322 patients who had achieved a CCyR by the European Study Group on Interferon in CML [97]. So, in the analysis of long-term survival, patients in either IRIS treatment arm could be considered as CCyR or nonresponders. For

responders, survival was modeled based on the survival distribution from the European Study Group. For nonresponders, survival data were available from the Italian Study Group trial.

Modeling survival over time usually requires some type of extrapolation method, such as a Markov chain model, where each year the surviving cohort is assumed to have an additional year of survival or to have expired during the year (death is an absorbing state in these models). This approach has some unusual characteristics, including generating a distributional tail of patients who survive out to very advanced age. Instead of using this approach, this study used a unique approach of modeling disease and treatment response as a function of increased risk of mortality on standard population survival models. Specifically, the researchers estimated log hazard ratios of the increased risk of death among patients with and without CCyRs compared with an age-matched and gender-matched cohort from the general population. They used the log hazard ratios and their standard errors to simulate survival curves for patients achieving and not achieving a CCyR at two years of follow-up, conditional on the patients being alive at two years. This approach was so unique that the survival estimation methods were published in a separate article concurrently with the economic analysis [98].

Advanced Phases of CML

During advanced phases of CML, the researchers assigned simulated patients in both treatment groups the same distributions for time in accelerated phase and blast crisis and the same distributions of utility weights and estimates of resource use using the published literature. Thus, there was assumed to be no additional benefit from the study therapy once disease had progressed.

Resource Use

Estimates of monthly counts of resource use, including medication use, physician visits, and hospital utilization, were based on data collected in the case report forms for patients in the IRIS study for patients receiving first-line therapy. The analysts assumed conservatively that patients receiving hydroxyurea experienced the same level of resource use as patients receiving imatinib. In the advanced phases of CML, estimates of resource use were based on pooled data from patients in both treatment groups in either the accelerated phase or blast crisis (again, all patients who progressed to this stage of disease were assumed to have the same costs). Resources were assigned the costs reported in Table 34.4 for the economic analysis.

Table 34.4 Resource costs for the economic analysis.

	Unit	Cost
Medication costs		
Imatinib	Cost per 100 mg	$19.68
Interferon-alpha	Cost per 1 MU	$12.24
Cytarabine	Cost per 100 mg	$ 5.74
Hydroxurea	Cost per 500 mg	$ 1.28
Chemotherapy		$6733
Outpatient visits		
Specialists		$78.93
Generalists		$78.93
Nurses		$34.73
Inpatient cost per day		
Chronic phase		$988.36
Accelerated phase		$1400.39
Blast phase		$1432.99

Utility Weights

As previously discussed, quality of life in an economic evaluation is based on an assessment of health utility. For the IRIS study, the EuroQol-5D (EQ-5D), a preference-based measure of health-related quality of life, was administered every three months to study patients. The EQ-5D has two components: a visual-analogue scale

(a health thermometer) and a set of five descriptive questions about health status that have categorical response scales. EQ-5D responses to the descriptive questions can be converted into standard community-weighted utility scores. In other words, patients on either treatment arm with the same response to the categorical questions would have the same utility weight assigned to their health status (there was no interaction of treatment arm and weighting).

Simulations

Once the model was developed, the analysts could calculate the costs and benefits of therapy for patients in either treatment group. Since all the inputs into the analysis had estimates of uncertainty, the modeling effect integrated this uncertainty into the analysis. For each population-level simulation run, the analysts simulated the outcomes for 1000 patients and computed mean estimates of costs, survival, and quality-adjusted survival for each treatment strategy. They varied the estimates for the parameters simultaneously according to their assigned distributions (each variable had a different distribution assigned to the data as appropriate), using means and standard errors. The results were reported as 95% confidence intervals (CIs) for the incremental cost-effectiveness ratios (ICERs), estimated using the percentile method, whereby the 26th and 975th rankings of the 1000 simulated ICERs were used as the 95% CI limits.

Sensitivity Analyses

While the analysts modeled uncertainty directly in the analysis, they had still made several critical assumptions that extended beyond the measurements of error in the model. Sensitivity analysis is an approach to addressing this issue, by questioning critical assumptions directly. The analysts conducted several sensitivity analyses to evaluate the impact of varying baseline estimates and assumptions. These included analyses which tested assumptions surrounding resource use, costs of medications, and outcomes.

Results

The model provided estimates of survival and costs for patients in each treatment arm. In the base-case analysis, patients receiving first-line therapy with imatinib were estimated to survive an average of 15.30 years compared to only 9.07 years for patients receiving first-line therapy with IFN+LDAC, an incremental gain of 6.23 years for imatinib. This is a very large treatment effect relative to other analyses of new pharmaceutical therapies. After adjusting for quality of life, the incremental gain was maintained at 5.85 QALYs (since having a disease like CML might impact your quality of life even if you are on effective therapy, the QALY estimates of benefits were less than the estimates of overall survival).

In terms of resource use, undiscounted lifetime costs were estimated to be $424 600 for patients receiving imatinib compared with $182 800 for patients receiving IFN+LDAC, a difference of $241 800. This cost difference included the difference in the cost of medications and the incremental additional years of treatment resulting from the additional survival of patients on imatinib.

After applying a 3% discount rate, the incremental gain in survival was 3.93 years and 3.89 QALYs. Discounted lifetime costs were $168 100 higher, on average, among patients receiving first-line therapy with imatinib. The net effect of discounting is to reduce the value of events occurring in the future, both on costs and survival.

Combining these estimates resulted in ICERs equal to $43 100 per life-year saved and $43 300 per QALY saved. Scatter plots of the ICERs from the 1000-patient analysis reveal a relatively low level of variability in the results and a high degree of correlation between incremental differences with regard to costs and survival. The results were remarkably consistent, with the upper limits of the 95% CIs for both cost-effectiveness

ratios (survival and QALY) less than $51 100 (Figure 34.3).

While the variability in individual parameters were considered in the scatter plot of Figure 34.3, there was still an analysis of variability of the different dimensions of the sensitivity analysis. Here, the results are presented as a tornado diagram, estimating the impact of each of the parameters of the sensitivity analysis on the overall study results (Figure 34.4).

This case study demonstrates an application of economic analysis to a new drug using data from the pivotal clinical trial to help justify the price for the therapy.

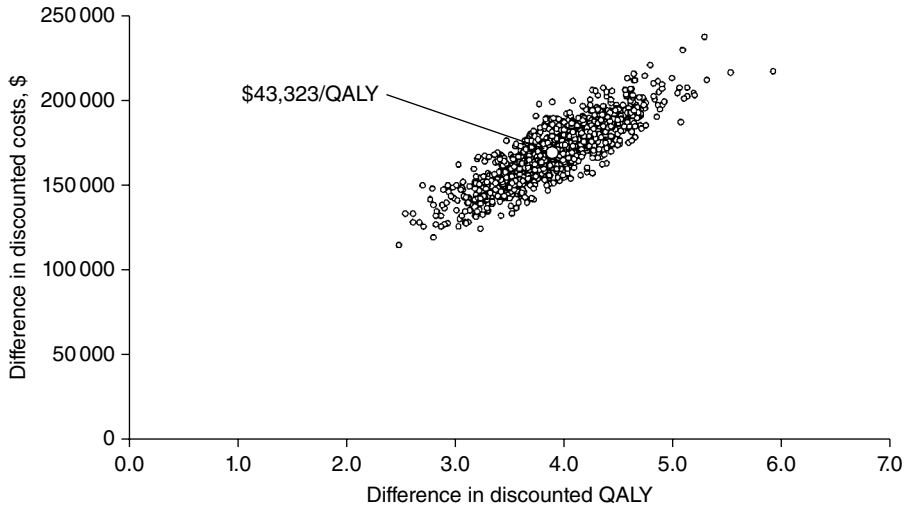


Figure 34.3 Bootstrap scatter plot of 1000 population-level simulations on an incremental cost-effectiveness plane [67].

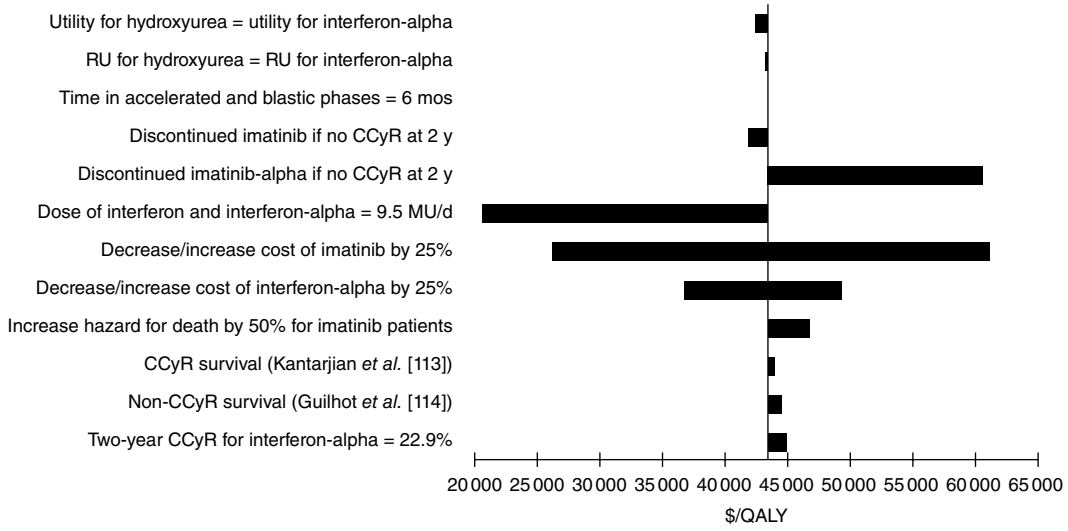


Figure 34.4 Tornado plot of the effects of multiple one-way sensitivity analyses [67].

This analysis was very controversial. First, the survival benefit in the analysis was entirely a modeling exercise, as a survival benefit had not been seen in the original IRIS study. Fortunately, this assumption of the relationship between CCyR and survival was later confirmed in a five-year follow-up of the original study [99]. Second, the economic analysis was based on the initial launch price of the therapy. However, the sponsor aggressively raised the price of therapy above the launch price over time, from $26 400 to over $120 000 [100]. Further, that price differs substantially by market, with costs in Canada less than 20% of the US costs in 2013 [101], down to a low of $2500 per year in India [102]. Patent expiry was expected to reduce the cost of imatinib significantly in the US [103], with a significant opportunity to reduce costs to a nominal level if the price comes down to the actual cost of active pharmaceutical ingredients [104].

The Future

Health economics helps to understand both the supply and demand for pharmaceutical products, while pharmacoeconomics provides insight into the value of products to patient, payers, and the marketplace. With all these elegant data, the next challenge for analysts and policy makers is to relate the results of the economic analysis to purchase decisions for pharmaceutical products. Should a patient, hospital, or payer (public or private) make a decision to approve payments for a therapy (for example, by adding the product to an approved drug list or formulary)?

Generally, if a product is thought to add clinical benefit and save money, it is an easy decision to add the therapy. These types of products are described as cost-saving or dominant (see Figure 34.2). For example, vaccines sometimes fall into this category, as do generic drugs (in comparison with brand-name products). If a product worsens clinical outcomes but raises cost, this is also usually an easy decision to not add the therapy. These types of products are described as dominated but are much less common. The major challenges in making formulary decisions generally relate to therapies that add clinical benefit at additional cost. This analysis requires further discussion on how this decision can be approached.

Outside healthcare environments, we make these types of decisions frequently. The new product on the shelf tastes better but costs more. Should we buy this product? The first question is a budget question – do we have the money to make this purchase? We have not said anything about prices but obviously, this is an important part of our consideration. If the new product is enormously more expensive (it is hand crafted in small batches and infused with gold and sold in a crystal decanter), we may not have the money to afford the new product and so the question becomes moot. This is a budget constraint. This constraint does not have to be so extreme; we can make a budget of $100 for a grocery list, and hold ourselves to meeting our budget in our shopping trip. With this budget, even a modestly priced new item may not meet our budget constraint.

Of course, we may decide we have some room in our budget to increase spending at the supermarket. For simplicity, let's assume we have two choices to consider for our increased spending: products A and B. Product A is the tastier version of one of our shopping staples, while product B is a new item that a friend recommended. Product A costs $5.00 more than our usual item, and product B costs $5.00, so our budget would now be $105 if either is added to the shopping list (a 5% increase in cost). How would you choose between these two? This is a cost-effectiveness question. Given the increase in cost, which product would provide more value to you as a consumer-enhanced flavor from A or the novelty of B? We make these types of decisions all the time, and the answer greatly depends

on individual taste preferences. So, in this simple case, cost is transparent and value is based on individual preferences.

Now, back to the pharmaceutical market. Budget constraints are built into healthcare spending. National health insurance programs often have a fixed allocation from government for annual spending. In private health insurance markets, health insurers estimate premiums for the coming year before selling policies during open-enrollment season, or as much as 18 months in advance of actual spending. So, the introduction of a new product that adds cost can face real budget constraints depending on the potential magnitude of the spending increase. Sofosbuvir had a list price of $84 000 in the US and a potential market size of 3.2 million people when it was first launched [105]. This would require a staggering budget of $269 billion to treat everyone with the infection. So, while the clinical potential of this therapy was tremendous, the budget constraint resulted in policies to limit its adoption.

As with the shopping example, budget constraints can be absolute. For example, the Medicaid program is jointly funded by federal and state governments, and states are not allowed to run budget deficits by law. Thus, if increased spending on a new therapy would require an increase in outlay by the Medicaid program, states may be forced to not offer the therapy, or to cut back in other areas of spending to stay within their budget. Going forward, if states wanted to add to their Medicaid spending, they would need to raise revenues (taxes) to support this increased spending, or find other parts of the budget to cut (education, for example).

Again, as with the shopping example, budget constraints may not be absolute. Imagine that a health insurance company calculated its premiums to incorporate new spending on drugs to be introduced in the coming year. Thinking about the budget from the perspective of premiums is interesting. We have called this approach the "benefit pool" perspective [106]. It suggests that we can calculate how much additional premium we would all have to contribute for us to have an increase in our pharmaceutical budget. The budget model looks at the issue from the insurer perspective, while the benefit pool perspective looks at the same issue from the perspective of everyone buying insurance (or paying taxes). With an increase in premium, we would still have a budget constraint, but one that allows for growth in pharmacy spending.

With the additional resources, we would need to develop a process for increasing our pharmacy spending. In our cash payment shopping model, the willingness to allocate additional resources to our budget was based on consumers' individual perceptions of value. In healthcare, we do not pay for our medicines directly, so the organization administering the benefit pool (which could be a public or private payer) needs to make this determination. Here, cost-effectiveness analysis can be used to assess the relative value of additional investments in different therapies. From here, one can consider the relative value across products and fund the product that is the most economically attractive (the lowest cost-effectiveness ratio) first. In this way, we will insure that incremental spending is for the product delivering the most value. You can continue adding therapies in this way until all of your resources are allocated [107].

Another way to use a cost-effectiveness ratio is to set a criterion of what represents good value for money, or what is "economically attractive." In the US, dialysis care has long been used to provide a benchmark of good value for money. Dialysis was added to the Medicare program in 1972, after consideration of the cost of care for patients with end-stage renal disease [108]. As a result, we have an example of a clinical program where Congress made an explicit decision to add a benefit to Medicare, one that added cost but that also extended life expectancy for beneficiaries who need the service. Since patients on dialysis can cost $50 000 per

annum, the benchmark for value as reflected in Congressional approval of this service was seen to be $50 000 per year of life gained. Since dialysis requires treatment three times a week for several hours at a time, the benchmark was thought to be even higher when considering quality-adjusted survival.

Does that mean that we can add therapies to the formulary that offer good value? The answer is, "it depends." Again, while cost-effectiveness analysis does a good job assessing the relative value of different therapies, the ratio itself is not tied to a budget impact or premium. In other words, a product that is not good value for money for a rare condition would have a relatively modest budget impact, while a drug that is good value for money for a common condition could have a significant impact on budgets. The value of a product can also change by indication. For example, since patients with known heart disease have higher risk for cardiovascular events than patients without heart disease, secondary prevention can provide more value for money than primary prevention [109]. To date, efforts such as pricing by indication have been challenging to implement.

Rather than providing an absolute recommendation, the UK has an implicit relative framework for value, with products that have a lower cost-effectiveness ratio more likely to be recommended by NICE [110].

Returning to the benefit pool perspective can be another way of looking at this question. This analysis looks at the impact on the health insurance premium resulting from the addition of a new product to the formulary. For example, the addition of PCSK-9 inhibitors (used to lower cholesterol) priced at more than $14 000 annually per patient was calculated to add $140 to the premium for everyone in the insurance pool under modest adoption assumptions [106]. This perspective can be generalized to the consideration of specialty pharmaceutical products more broadly (Figure 34.5). Here, we can see that health insurance premiums increase $250 for every 0.25% of the population that receives a $100 000 drug for any indication [23]. The relationship of access to innovation and affordability is an area of ongoing debate [111].

Finally, we have the quadrant where the products save money but at the expense of worse clinical outcomes. Actually, the use of therapies that meet this criterion is relatively common, as long as the amount of money saved is large relative to the loss of health benefits (a cost-effectiveness ratio of savings related to benefits lost where a higher number is the most economically attractive). For example, amoxicillin is recommended

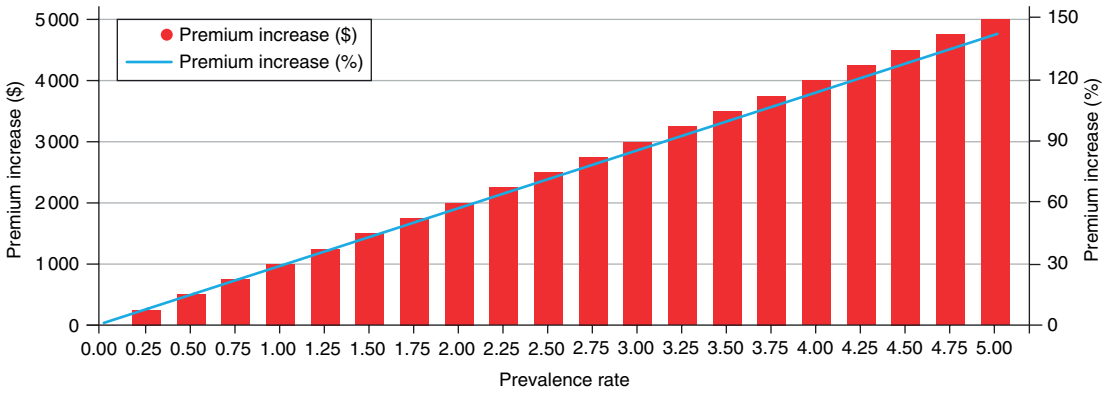


Figure 34.5 Specialty pharmaceuticals and premiums. *Source:* Hirsch *et al*. [24].

for first-line therapy for otitis media, despite the high level of resistance to this antibiotic [112]. This is because of the low cost of the therapy, and the low likelihood of significant complications of failure of this initial treatment.

Conclusion

The cost of pharmaceutical products is an important challenge for everyone involved in the healthcare value chain, including patients, physicians, payers, and government. While we all desire innovation in healthcare, it is not clear if it is possible to afford innovation at any price. We have outlined an expanded version of the study of the economics of pharmaceuticals to begin to flesh out a fuller discussion of this fascinating and complex subject.

As physicians are asked simultaneously to represent their patients' interests while delivering clinical services with parsimony, and as reimbursement for medical services becomes more centralized in the United States and other countries, decision makers must turn for assistance to collaborative efforts of epidemiologists and economists in the assessment of new therapeutic agents. Through a merger of epidemiology and economics, better information can be provided to the greatest number of decision makers, and limited resources can be used most effectively for the health of the public.

Acknowledgments

The author dedicates this chapter to John M. Eisenberg, MD (1946–2002), mentor, friend, and fellow author in previous editions of this work. The author also acknowledges Henry Glick PhD, Harris Koffer Pharm D, Dan Polsky PhD, and Shelby Reed PhD for contributions to previous versions of this chapter.

References

1 DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016; **47**: 20–33.
2 Mukherjee S. Is $475,000 Too High for Novartis's 'Historic' Cancer Gene Therapy? http://fortune.com/2017/08/31/novartis-kymriah-car-t-cms-price/ (accessed April 24, 2019).
3 2015 Comparative Price Report: Variation in Medical and Hospital Prices by Country. International Federation of Health Plans. https://books.google.com/books/about/International_Federation_of_Health_Plans.html?id=eASZtAEACAAJ (accessed April 24, 2019).
4 Memorial Sloan Kettering Cancer Center. Price and Value of Cancer Drug. www.mskcc.org/research-areas/programs-centers/health-policy-outcomes/cost-drugs (accessed April 24, 2019).
5 Centers for Medicare & Medicaid Services. NHE 2015. www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html (accessed April 24, 2019)
6 Outlook for Global Medicines Through 2021: Balancing Cost and Value. Quintiles IMS Institute. https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/global-outlook-for-medicines-through-2021.pdf?la=en&hash=6EA26BACA0F1D81EA93A74C50FF60214044C1DAB (accessed April 24, 2019).
7 Arrow KJ. Uncertainty and the welfare economics of medical care. *AER* 1963; **53**(5): 141–9.

8 Pauly MV. The economics of moral hazard: comment. *AER* 1968; **58**(3): 531–7.

9 Newhouse JP. A summary of the Rand Health Insurance Study. *Ann N Y Acad Sci* 1982; **387**: 111–14.

10 Nyman JA. Is 'moral hazard' inefficient? The policy implications of a new theory. *Health Affairs* 2004; **23**(5): 194–9.

11 Curtis LH, Law AW, Anstrom AJ, Schulman KA. The insurance effect on prescription drug expenditures among the elderly: findings from the 1997 Medical Expenditure Panel Survey. *Med Care* 2004; **42**(5): 439–46.

12 Rice TH. The impact of changing Medicare reimbursement rates on physician-induced demand. *Med Care* 1983; **21**(8): 803–15.

13 Debpuur C, Dalaba MA, Chatio S, Adjuik M, Akweongo P. An exploration of moral hazard behaviors under the national health insurance scheme in Northern Ghana: a qualitative study. *BMC Health Serv Res* 2015; **15**: 469.

14 Finkelstein A, Arrow KJ, Gruber J, Newhouse JP, Stiglitz JE. *Moral Hazard in Health Insurance*. New York: Columbia University Press, 2014.

15 Lakdawalla D, Sood N. Innovation and the welfare effects of public drug insurance. *J Public Econ* 2009; **93**(3-4): 541–8.

16 Havighurst CC, Richman BD. Distributive injustice(s) in American health care. *Law Contemp Prob* 2006; **69**(4): 7–82.

17 Schulman KA. The Supply-Side Effects of Moral Hazard on Drug Prices. *Health Management Policy and Innovation*, Volume 3, Issue 2, 2018.

18 Private investors expect a risk-adjusted return from their investments. If we assumed that the entire cost of drug development occurred at the start of the project, that the research program took six years, and that investors demanded a 30% annual return on their investment to account for the high-risk of failure in drug development, and the time value of money is 6%, the capitalized cost of development would be $288 million at the start of year 7.

19 Daniel Kahneman – Facts. www.nobelprize.org/prizes/economic-sciences/2002/kahneman/facts/ (accessed April 24, 2019).

20 Gaskin DJ, Weinfurt KP, Castel LD, *et al*. An exploration of relative health stock in advanced cancer patients. *Med Decis Making* 2004; **24**(6): 614–24.

21 Rasiel EB, Weinfurt KP, Schulman KA. Can prospect theory explain risk-seeking behavior by terminally ill patients? *Med Decis Making* 2005; **25**(6): 609–13.

22 Kahneman D. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux, 2013.

23 Ariely D. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: Harper Collins, 2010.

24 Hirsch BR, Balu S, Schulman KA. The impact of specialty pharmaceuticals as drivers of health care costs. *Health Affairs* 2014; **33**(10): 1714–20.

25 National Cancer Act of 1971. National Cancer Institute. www.cancer.gov/about-nci/legislative/history/national-cancer-act-1971 (accessed April 24, 2019).

26 Centers for Disease Control and Prevention. Women and Heart Disease Fact Sheet. www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_women_heart.htm (accessed April 24, 2019).

27 American Cancer Society Breast Cancer Facts and Figures, 2013–2014. www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2013-2014.pdf (accessed April 24, 2019).

28 Mukherjee S. *The Emperor of All Maladies: A Biography of Cancer*. New York: Scribner, 2010.

29 History of the Jimmy Fund. www.jimmyfund.org/about-us/about-the-jimmy-fund/history-of-the-jimmy-fund/ (accessed April 24, 2019).

30 Berkrot B. Biden Announces Crucial Piece of His Cancer Moonshot Initiative. Huffington

Post. www.huffingtonpost.com/entry/biden-announces-us-project-to-promote-cancer-data-sharing_us_5756fc88e4b0b60682df0d13 (accessed April 24, 2019).
31 Schulman KA. The economics of precision medicine. In: *Insights and Innovation in Neuromuscular Medicine*. Rochester: AANEM, 2016.
32 Jacobson IM, Gordon SC, Kowdley KV, *et al*. Sofosbuvir for hepatitis C genotype 2 or 3 in patients without treatment options. *N Engl J Med* 2013; **368**: 1867–77.
33 Liang TJ, Ghany MG. Current and future therapies for hepatitis C virus infection. *N Engl J Med* 2013; **368**: 1907–17.
34 Hepatitis C Online. Sofosbuvir. www.hepatitisc.uw.edu/page/treatment/drugs/sofosbuvir-drug (accessed April 24, 2019).
35 Liao JM, Fischer MA. Early patterns of sofosbuvir utilization by state Medicaid programs. *N Engl J Med* 2015; **373**(13): 1279–81.
36 Prepared by the Staffs of Ranking Member Ron Wyden and Committee Member Charles E. Grassley. The Price of Sovaldi and Its Impact on the U.S. Health Care System, DECEMBER 2015. US Senate, Committee of Finance. www.finance.senate.gov/imo/media/doc/1%20The%20Price%20of%20Sovaldi%20and%20Its%20Impact%20on%20the%20U.S.%20Health%20Care%20System%20(Full%20Report).pdf (accessed April 24, 2019).
37 National Institute for Health and Care Excellence. Sofosbuvir for Treating Chronic Hepatitis C. Technology Appraisal Guidance. www.nice.org.uk/guidance/ta330 (accessed April 24, 2019).
38 Southampton Health Technology Assessments Centre. Sofosbuvir for treating chronic hepatitis C. https://eprints.soton.ac.uk/414361/1/Sofosbuvir_for_treating_chronic_hepatitis_C.pdf (accessed April 24, 2019).
39 Pollack A. Sales of Sovaldi, New Gilead Hepatitis C Drug, Soar to $10.3 Billion. New York Times. www.nytimes.com/2015/02/04/business/sales-of-sovaldi-new-gilead-hepatitis-c-drug-soar-to-10-3-billion.html (accessed April 24, 2019).
40 Sagonowsky E. AbbVie's new pan-genotypic hepatitis C drug Mavyret deeply underprices the competition. www.fiercepharma.com/pharma/abbvie-s-new-pan-genotypic-hep-c-drug-mavyret-undercuts-competition (accessed April 24, 2019).
41 Gilead Sciences Announces Fourth Quarter and Full Year 2016 Financial Results. http://investors.gilead.com/static-files/33588c5a-7f81-437a-b35a-379514d49eff (accessed April 24, 2019).
42 Bombardier C, Eisenberg J. Looking into the crystal ball: can we estimate the lifetime cost of rheumatoid arthritis? *J Rheumatol* 1985; **12**(2): 201–4.
43 Drummond MF. *Principles of Economic Appraisal in Health Care*. New York: Oxford University Press, 1980.
44 Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Evaluation of Health Care Programs*, 3rd edn. New York: Oxford Medical Publications, 2005.
45 Drummond MF, Davies L. Economic analysis alongside clinical trials: revisiting the methodological issues. *Int J Technol Assess Health Care* 1991; **7**(4): 561–73.
46 Detsky AS, Naglie IG. A clinician's guide to cost effectiveness analysis. *Ann Intern Med* 1990; **113**(2): 147–54.
47 Eddy DM. Cost-effectiveness analysis: is it up to the task? *JAMA* 1992; **267**(24): 3342–8.
48 Eisenberg JM, Koffer H, Finkler SA. Economic analysis of a new drug: potential savings in hospital operating costs from the use of a once-daily regimen of a parenteral cephalosporin. *Rev Infect Dis* 1984; **6**(Suppl 4): S909.
49 Eisenberg JM, Koffer H, Glick HA, *et al*. What is the cost of nephrotoxicity associated with aminoglycosides? *Ann Intern Med* 1987; **107**(6): 900–9.

50 Eisenberg JM. Clinical economics: a guide to the economic analysis of clinical practices. *JAMA* 1989; **262**(20): 2879–86.

51 Glick HA, Connell ML, Koffer H, Eisenberg JM, Kelley MA. The economic impact of prophylaxis against venous thromboembolism in patients undergoing total hip replacement. *Clin Res* 1988; **36**: A337.

52 Warner KE, Luce BR. *Cost Benefit and Cost Effectiveness Analysis in Health Care: Principles, Practice and Potential*. Ann Arbor, MI: Health Administration Press, 1982.

53 Weinstein MC. Economic assessments of medical practices and technologies. *Med Decis Making* 1981; **1**(4): 309.

54 Weinstein MC, Stason WB. Cost-effectiveness of coronary artery bypass surgery. *Circulation* 1982; **66**(5): 56–68.

55 Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *Int J Technol Assess Health Care* 1990; **6**(1): 93–103.

56 George B, Harris A, Mitchell A. Cost-effectiveness analysis and the consistency of decision making: evidence from pharmaceutical reimbursement in Australia (1991 to 1996). *Pharmacoeconomics* 2001; **19**(11): 1103–9.

57 Menon D. Health technology assessment: the journey continues. *CMAJ* 2015; **187**(1): E19–20.

58 Sorenson C, Chalkidou K. Reflections on the evolution of health technology assessment in Europe. *Health Econ Policy Law* 2012; **7**(1): 25–45.

59 Gulácsi L, Rotar AM, Niewada M, *et al*. Health technology assessment in Poland, the Czech Republic, Hungary, Romania and Bulgaria. *Eur J Health Econ* 2014; **15**(Suppl 1): S13–25.

60 Drummond M, Neumann P, Jönsson B, *et al*. Can we reliably benchmark health technology assessment organizations? *Int J Technol Assess Health Care* 2012; **28**(2): 159–65.

61 Raftery J, Powell J. Health technology assessment in the UK. *Lancet* 2013; **382**(9900): 1278–85.

62 Institution for Quality and Efficiency in Health Care. Responsibilities and objectives of IQWiG. www.iqwig.de/en/about-us/responsibilities-and-objectives-of-iqwig.2946.html (accessed April 24, 2019).

63 Chahal HS, Marseille EA, Tice JA, *et al*. Cost-effectiveness of early treatment of hepatitis C virus genotype 1 by stage of liver fibrosis in a US treatment-naive population. *JAMA Intern Med* 2016; **176**(1): 65–73.

64 Ramsey S, Willke R, Briggs A, *et al*. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* 2005; **8**(5): 521–33.

65 Mark DB, Hlatky MA, Califf RM, *et al*. Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *N Engl J Med* 1995; **332**: 1418–24.

66 Reed SD, Radeva JI, Glendenning GA, Saad F, Schulman KA. Cost-effectiveness of zoledronic acid for the prevention of skeletal complications in patients with prostate cancer. *J Urol* 2004; **171**(4): 1537–42.

67 Reed SD, Anstrom KJ, Ludmer JA, Glendenning GA, Schulman KA. Cost-effectiveness of imatinib versus interferon-alpha plus low-dose cytarabine for patients with newly diagnosed chronic-phase chronic myeloid leukemia. *Cancer* 2004; **101**(11): 2574–83.

68 Reed SD, Friedman JY, Velazquez EJ, Gnanasakthy A, Califf RM, Schulman KA. Multinational economic evaluation of valsartan in patients with chronic heart failure: results from the Valsartan Heart Failure Trial (Val-HeFT). *Am Heart J* 2004; **148**(1): 122–8.

69 Reed SD, Radeva JI, Weinfurt KP, *et al*. Resource use, costs, and quality of life among patients in the multinational Valsartan in

Acute Myocardial Infarction Trial. *Am Heart J* 2005; **150**(2): 323–9.

70 Reed SD, Anstrom KJ, Bakhai A, *et al*. Conducting economic evaluations alongside multinational clinical trials: toward a research consensus. *Am Heart J* 2005; **149**(3): 434–43.

71 Hlatky MA, Owens DK, Sanders GD. Cost-effectiveness as an outcome in randomized clinical trials. *Clin Trials* 2006; **3**(6): 543–51.

72 Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; **290**(12): 1624–32.

73 Mendell JR, MD, Rodino-Klapac LR, Sahenk Z, *et al*. Eteplirsen for the treatment of Duchenne muscular dystrophy. *Ann Neurol* 2013; **74**(5): 637–64.

74 Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. *Med Decis Making* 2000; **20**(3): 332–42.

75 Towse A, Pritchard C. Does NICE have a threshold? An external view. In: Towse A, Pritchard C, Devlin N, eds. *Cost-Effectiveness Thresholds: Economic and Ethical Issues*. London: Kings Fund/Office for Health Economics, 2002.

76 Finkler SA. The distinction between cost and charges. *Ann Intern Med* 1982; **96**(1): 102–9.

77 Bridges JM, Jacobs P. Obtaining estimates of marginal cost by DRG. *Healthcare Financ Manage* 1986; **40**(10): 40–6.

78 Hsiao WC, Braun P, Becker ER, *et al*. An overview of the development and refinement of the resource-based relative value scale: the foundation for reform of U.S. physician payment. *Med Care* 1992; **30**(suppl): NS1–NS12.

79 Granneman TW, Brown RS, Pauly MV. Estimating hospital costs. *J Health Econ* 1986; **5**(2): 107–27.

80 Reed SD, Friedman JY, Gnanasakthy A, Schulman KA. Comparison of hospital costing methods in an economic evaluation of a multinational clinical trial. *Int J Technol Assess Health Care* 2003; **19**(2): 396–406.

81 Landefeld JS, Seskin EP. The economic value of life: linking theory to practice. *Am J Public Health* 1982; **72**(6): 555–65.

82 Rice DP, Hodgson TA. The value of human life revisited. *Am J Public Health* 1982; **72**(6): 536–8.

83 Thompson MS, Read JL, Liang M. Feasibility of willingness-to-pay measurement in chronic arthritis. *Med Decis Making* 1984; **4**(2): 195–215.

84 Mishan EJ. *Cost–Benefit Analysis*, 3rd edn. London: George Allen & Unwin, 1992.

85 Shelling TC, Chase SB, eds. *The life you save may be your own. In: Problems in Public Expenditure Analysis*. Washington, DC: Brookings Institution, 1968.

86 Lieu TA, Finkler LJ, Sorel ME, Black SB, Shinefield HR. Cost-effectiveness of varicella serotesting versus presumptive vaccination of school-age children and adolescents. *Pediatrics* 1995; **95**(5): 632–8.

87 Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA* 2003; **289**(23): 3135–44.

88 Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. *JAMA* 2003; **290**(18): 2443–54.

89 Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.

90 Strosberg MA, Wiener JM, Baker R, Fein IA, eds. *Rationing America's Medical Care: The Oregon Plan and Beyond*. Washington, DC: Brookings Institution, 1992.

91 Sulmasy DP. Physicians, cost-control, and ethics. *Ann Intern Med* 1992; **116**: 920–6.

92 An X, Tiwari AK, Sun Y, Ding PR, Ashby CR Jr, Chen ZS. BCR-ABL tyrosine kinase

inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leuk Res* 2010; **34**(10): 1255–68.

93 National Cancer Institute. How Imatinib Transformed Leukemia Treatment and Cancer Research. www.cancer.gov/research/progress/discovery/gleevec (accessed April 24, 2019).

94 Vasella D, Slater R. *Magic Cancer Bullet: How a Tiny Orange Pill is Rewriting Medical History*. New York: Harper Business, 2003.

95 O'Brien SG, Guilhot F, Larson RA, *et al*. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukaemia. *N Engl J Med* 2003; **348**: 994–1004.

96 Baccarani M, Rosti G, de Vivo A, *et al*. A randomized study of interferon-alpha versus interferon-alpha and low-dose arabinosyl cytosine in chronic myeloid leukemia. *Blood* 2002; **99**(5): 1527–35.

97 Bonifazi F, de Vivo A, Rosti G, *et al*. Chronic myeloid leukemia and interferon-alpha: a study of complete cytogenetic responders. *Blood* 2001; **98**(10): 3074–81.

98 Anstrom KJ, Reed SD, Allen AS, Glendenning GA, Schulman KA. Long-term survival estimates for imatinib versus interferon-α plus low-dose cytarabine for patients with newly diagnosed chronic-phase chronic myeloid leukemia. *Cancer* 2004; **101**(11): 2584–92.

99 Reed SD, Anstrom KJ, Li Y, Schulman KA. Updated estimates of survival and cost effectiveness for imatinib versus interferon-alpha plus low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukaemia. *Pharmacoeconomics* 2008; **26**(5): 435–46.

100 Johnson CY. This Drug is Defying a Rare Form of Leukemia – And It Keeps Getting Pricier. www.washingtonpost.com/business/this-drug-is-defying-a-rare-form-of-leukemia--and-it-keeps-getting-pricier/2016/03/09/4fff8102-c571-11e5-a4aa-f25866ba0dc6_story.html?utm_term=.cafdc08f3ac7 (accessed April 24, 2019).

101 Average prices of Gleevec in selected countries in 2013 (in U.S. dollars). www.statista.com/statistics/312011/prices-of-gleevec-by-country/ (accessed April 24, 2019).

102 Bollyky T. Why Chemotherapy that Costs $70,000 in the US Costs $2,500 in India. www.theatlantic.com/health/archive/2013/04/why-chemotherapy-that-costs-70-000-in-the-us-costs-2-500-in-india/274847/ (accessed April 24, 2019).

103 Padula WV, Larson RA. Changing the cost of care for chronic myeloid leukemia: the availability of generic imatinib in the USA and the EU. *Ann Hematol* 2015; **94**(Suppl 2): 249–57.

104 Hill A, Gotham D, Fortunak J, *et al*. Target prices for mass production of tyrosine kinase inhibitors for global cancer treatment. *BMJ Open* 2016; **6**: e009586.

105 Hepatitis C. www.cdc.gov/hepatitis/statistics/2016surveillance/commentary.htm (accessed April 24, 2019).

106 Schulman KA, Reed SD. The economics of PCSK-9 Inhibitors. Research Letter. *Am Heart J* 2017; **189**: 1–2.

107 Klevit HD, Bates AC, Castanares T, Kirk EP, Sipes-Metzler PR, Wopat R. Prioritization of health care services. A progress report by the Oregon Health Services Commission. *Arch Intern Med* 1991; **151**(5): 912–16.

108 Klarman HE, Francis JO'S, Rosenthal GD. Cost effectiveness analysis applied to the treatment of chronic renal disease. *Med Care* 1968*;* **6***(*1*):* 48–54.

109 Goldman L, Weinstein MC, Goldman PA, Williams LW. Cost-effectiveness of HMG-CoA reductase inhibition for primary and

secondary prevention of coronary heart disease. *JAMA* 1991; **265**(9): 1145–51.

110 Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. The influence of cost-effectiveness and other factors on NICE decisions. *Health Econ* 2015; **24**(10): 1256–71.

111 Mark D, Schulman KA. PCSK9 Inhibitors and the choice between innovation, efficiency, and affordability. *JAMA* 2017; **318**(8): 711–12.

112 Harmes KM, Blackwood A, Burrow HL, Cooke JM, Harrison V, Passamani PP. *Otitis media: diagnosis and treatment Am Fam Physician* 2013; **88**(7): 435–40.

113 Kantarjian HM, O'Brien S, Cortes JE, *et al*. Complete cytogenetic and molecular responses to interferon-alpha-based therapy for chronic myelogenous leukemia are associated with excellent long-term prognosis. *Cancer* 2003; **97**: 1033–41.

114 Guilhot F, Chastang C, Michallet M, *et al*. Interferon alfa-2b combined with cytarabine versus interferon alone in chronic myelogenous leukemia. French Chronic Myeloid Leukemia Study Group. *N Engl J Med* 1997; **337**: 223–9.

35

Benefit–Risk Assessments of Medical Treatments

Bennett Levitan[1], Rachael L. DiSantostefano[1], and Scott Evans[2]

[1] *Department of Epidemiology, Janssen Research & Development, Titusville, NJ, USA*
[2] *Biostatistics Center, The George Washington University, Rockville, MD, USA*

Consider the challenges faced by a health authority reviewing a novel anticoagulant for people with atrial fibrillation. The drug reduces ischemic strokes and myocardial infarctions, but also causes intracranial hemorrhage, gastric bleeding, and increased propensity for hematomas. What chance of each type of bleed is too much for a given degree of ischemic benefit? Does this trade-off differ for different groups of patients? Does the nature of this trade-off differ between physicians and patients? How much uncertainty in the rates of these outcomes is acceptable? How can these trade-offs be assessed in observational databases where the outcome definitions vary and detailed information on transfusions and hemoglobin changes is not always available? Addressing these questions falls under the domain of benefit–risk (B–R) assessment.

Assessing the B–R balance of medical treatments has always been an integral part of drug development, regulatory, and public health decisions. However, methodology for B–R has advanced considerably in the last decade, and health authorities worldwide have seen growth and maturation of B–R policies relating to regulatory approval and postapproval decisions. Starting from a report on a 2006 Institute of Medicine public workshop which noted the need for improved consistency, methodology and communication of B–R assessment [1], the convergence of regulatory science and policy, patient engagement, decision analysis, and health economics has transformed the B–R field. For example, the Food and Drug Administration (FDA) has implemented a B–R framework to facilitate B–R assessment and communication in its reviews [2] and is developing means to include the patient perspective into B–R [3–7]. The European Medicines Agency revised its 80-day assessment reports to require rapporteurs to document the value judgments behind their B–R assessments and incorporate a detailed tabular summary of benefit and risk data [8,9]. The International Conference on Harmonization (ICH) revised guidance documents that now require pharmaceutical companies to use a structured approach to B–R in new drug applications/marketing authorization applications and in postapproval periodic benefit–risk evaluation reports (PBRERs) [10,11]. In parallel with these regulatory changes, numerous initiatives have been led by pharmaceutical and device trade organizations, patient advocacy groups,

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

public–private partnerships, and academic groups. Their efforts have resulted in new qualitative and quantitative approaches to B–R, means to assess patient B–R preferences and incorporate them in B–R assessments, and compendia of approaches and recommendations [12–18,29]. These advances have driven the addition of this inaugural chapter in B–R assessment.

While definitions vary, benefit–risk is generally defined as weighing the key benefits of a treatment against its key harms. The term "risk" is ambiguous in the field of benefit–risk, and may refer to the general nature of the harmful effect, its frequency, its severity or a vague combination of these concepts [19,20]. To lessen this ambiguity, we will generally use the term "harm" or "unfavorable effect" as the analogs for "benefit" or "favorable effect," though we retain the phrase "benefit–risk" due to its ubiquity. The goals for B–R assessments vary – go/no-go decisions by a company, reviews by institutional review boards and data monitoring committees, patient group advocacy decisions, regulatory reviews by a health authority, and reconsideration of a treatment after gaining new information post approval. Benefit–risk is also a critical component of point-of-care decision making by a patient and physician.

The goals for this chapter are to introduce the key clinical and methodological challenges in B–R assessment, then summarize current and promising approaches to addressing these challenges. The focus of this book is pharmacoepidemiology, but because medical treatment B–R as a science has much of its origin and current applications in development and regulatory approval, this chapter touches on both regulatory development and postapproval settings. Our examples will primarily be in the cardiovascular domain – for the treatment of atrial fibrillation or acute coronary syndrome (i.e., myocardial infarction or unstable angina).

Clinical Problems to be Addressed by Pharmacoepidemiologic Research

Systematic Approach to B–R Assessment

While many B–R assessments are ultimately based on a qualitative interpretation of quantitative data [2,18], reaching the point where this qualitative interpretation is possible can be surprisingly challenging. Benefit–risk assessment can involve many outcomes, multiple conditions under which these outcomes are assessed, many data sources, multiple regimens, numerous comparators, different treatment paradigms, and numerous other considerations. There are also often several stakeholders whose perspectives on the importance of benefits and harms outcomes may need to be considered. All this information needs to be integrated into a transparent and defensible B–R assessment. There are also numerous quantitative models available for assessing B–R [16,21,22]. Because of all these complexities, different stakeholders use widely varying approaches to B–R assessment, with considerable differences in the depth, transparency, and clarity of the assessment. This has resulted in considerable inconsistency in how, for example, drug or device companies and regulators conduct and communicate a B–R assessment. What has emerged in recent years is the need for structured, framework approaches to rationally and defensibly frame, conduct, and communicate a B–R assessment [2,7,10,11,18,23]. We describe several of these frameworks and how they can be used.

Incorporating the Patient Perspective

Traditionally, the judgment calls required in design and conduct of clinical trials, registries, long-term extension studies, and the B–R assessments on their results have been the province of

physicians and regulators. Regulatory agencies, patient advocacy groups, and industry have increasingly been advocating for patient engagement: patients partnering with the drug/device development company, researchers, health authorities and payers in the development, review, reimbursement and public health (e.g., vaccines) decisions for their treatments [5,24–28]. Chapter 42 addresses the topic of patient engagement in detail. Here, we consider patient-focused benefit–risk, which can be regarded as the components of patient engagement involving the incorporation of the patient perspective on the nature of the illness, medical need, outcomes and trade-offs into a B–R assessment. While the idea of patient-focused B–R is well accepted, the challenge is to determine what information should be obtained from patients, how it should be obtained, and how it can be used.

There are numerous techniques for assessing the patient perspective, ranging from qualitative focus groups and structured interviews through rigorously quantitative conjoint analysis or discrete choice analysis preference surveys [16,19,29,30]. Patient preference studies and surveys have become particularly important [3,4,15,19,31–33]. Identifying the appropriate patient sample and methodology, establishing trust between the stakeholders involved in these surveys, and the application of the results to clinical trials or observational studies are part art and part science. We review key elements of patient preference studies and discuss several means by which the results can be incorporated into a B–R assessment.

Methodologic Problems to be Addressed by Pharmacoepidemiologic Research

Regulators, patients, and physicians need to understand the evidence regarding the benefits and harms of available therapies to make optimal therapeutic decisions. Methodologies for gathering, synthesizing, and communicating the B–R assessment based on available data are the key tools for optimal decision making, in both the regulatory and clinical realms.

Identifying Appropriate Data for Benefit–Risk Assessment

Data for B–R assessment come from a variety of sources including randomized controlled trials, spontaneous reports, and observational data sources (see Part III Sources of Data for Pharmacoepidemiologic Studies). These sources often define outcomes differently, measure effects over different time periods, have important differences between populations, and use different comparators. Even with similar data sources, studies can demonstrate dissimilar results. Some studies might find an association between a treatment and adverse event, while others may find a protective effect or no effect. There has been much discussion about discordant results between observational studies and randomized controlled trials and potential explanations [34–36]. A well-known discrepancy is the difference between the Nurses' Health Study, an observational cohort, and the Women's Health Initiative, a large randomized trial, on the relationship between hormone replacement therapy and cardiovascular and breast cancer outcomes. The populations studied differed on important confounders and effect modifiers [37–39].

The reliability in identifying and characterizing the important benefits and harms must be carefully evaluated through assessment of confounding, bias, study design, generalizability and ability to pool the data. The principles for choosing among the available data sources in pharmacoepidemiology are discussed in Chapter 17. The validity of exposure and diagnostic data is further discussed in Chapter 37. Design-based and analytic approaches to adjust for confounding and bias in observational studies are discussed in Chapter 43. The considerations in

these chapters for pharmacoepidemiology also apply to B–R assessment.

Integrating Benefits and Risks

Even with high-quality data, synthesizing the evidence in a structured manner can be challenging. For example, a typical psoriasis indication can easily have over 10 efficacy outcomes and 15 important safety outcomes. Some of these outcomes may favor one treatment, while others may favor the other. Even some "benefits" may favor the comparator, and some "harms" may favor the study treatment. Although comparisons may reach statistical significance, in benefit–risk, there are often several important outcomes for which formal statistical testing is not conducted, for the most part due to small numbers of events and insufficient power to detect differences. In cases like this, rendering a B–R assessment is not trivial. Approaches that integrate benefits and harms should be sufficiently comprehensive to account for the frequency, clinical impact, and uncertainty of potentially many benefits and harms considered under multiple conditions. Further, they should account for the context of the overall disease state, currently available treatments, and unmet needs of the population of interest. This challenge of integrating available evidence is complicated by the need to evaluate the effects upon important subgroups of patients, for example, children, the elderly or immune-compromised individuals, where the B–R may be more or less favorable. Different subgroups may also have differences in preferences and value judgments that impact the B–R balance.

Communicating B–R Assessment

Given the many outcomes, complex statistics or epidemiology, and clinical judgments or formal preference studies, communicating a B–R assessment in a clear and cogent manner can be a formidable challenge (see also Chapter 39 on risk communication). This challenge is compounded by the typically very heterogeneous background of the audience (regulatory reviewers, patient groups, clinicians) that might utilize this information in a B–R assessment.

Currently Available Solutions

Structured Approaches to B–R: B–R Frameworks

Perhaps the most significant advance in regulatory and industry approaches to B–R in the last decade has been the introduction of B–R frameworks, a set of principles, processes, and tools to guide decision makers in selecting, organizing, analyzing, and communicating evidence relevant to B–R decisions [18]. B–R frameworks lead to far greater transparency, consistency, and discipline in B–R decision making. They encourage discussions that force decision makers to be explicit about their assumptions and requirements and, as will be discussed later, they assist not only in the act of decision making, but in the act of communicating the decision.

There are several well-known B–R frameworks: the FDA B–R framework [2], the Pharmaceutical Research and Manufacturers of America Benefit-Risk Action Team (PhRMA BRAT) framework [13,14], the Multicriteria Decision Analysis (MCDA) framework suggested by the European Medicines Agency (EMA) for complex B–R decisions [40–42], and the framework embedded in the International Council for Harmonisation's (ICH) PBRER template and recent update to the ICH Clinical Overview template's B–R section [10,11]. There are also applications of the BRAT framework for nonprescription drugs [43] and a documentation system for regulatory agencies based on both BRAT and the ICH Clinical Overview update [17]. While the frameworks differ in focus and methodology, they share many steps

Table 35.1 Steps in the BRAT benefit–risk framework.

Step	Definition
Define decision context	Summarize the nature of disease, medical need for treatment, disease and treatment epidemiology, study treatment, dose/formulation, indication(s), patient population, critical subgroups, comparator(s), time horizon for outcomes, relevant decision-making bodies
Identify and define outcomes	Identify and define all important outcomes, define a preliminary set of measures for each outcome, document rationale for outcomes to be included and excluded
Identify and summarize source data	Determine and document all data sources, extract raw data, summarize over data sources, assemble effects table
Customize the framework	Modify the outcome list and their definitions based on review of the data and clinical expertise. May include tuning of outcomes not considered relevant to a particular B–R assessment or stakeholder group
Assess importance of outcomes – value judgments and patient preferences	If applicable, assess outcome clinical impact or weight from the perspective of patients, decision makers or other stakeholders
Integrated B–R assessment: analysis and visualization	Summarize data into tabular and graphical displays (e.g., effects table) to aid interpretation, identify and fill any information gaps, interpret summary information, potentially conduct quantitative B–R analyses and sensitivity analyses to assess the impact of uncertainty on clinical or preference data
Expert judgment and communication	Render and communicate a decision

similar to those in the BRAT framework presented in Table 35.1.

Given the ubiquity of B–R frameworks and their incorporation into the regulatory guidance for submissions and PBRERs, we structure this section based on these framework steps.

Define Decision Context

The main role of the decision context is to ensure all decision makers are aligned on key background elements of the assessment. Characterizing the severity of the disease and the medical need is particularly important, as the greater the severity or medical need, the more allowances regulators and other decision makers typically make for treatment-related harms [7]. The severity of the illness can reflect severity either untreated or with current standard of care therapy, whichever is most appropriate for the condition. Medical need typically reflects the B–R profiles or limitations of existing treatments. These background elements are also an important avenue for incorporating the patient perspective into B–R [6]. In observational B–R assessments, the context forces upfront agreement on an appropriate comparator, dose and population, the choice of which is often not straightforward in postapproval contexts. For example, for a recently approved drug, the most appropriate comparator may be the one used in recent clinical trials, the standard of care, or comparators available in the extant datasets.

Examples of decision contexts for numerous disease can be found in the FDA's "Voice of the Patient" reports [25] and in recent drug approvals, searchable at www.accessdata.fda.gov/scripts/cder/daf/.

Identify and Define Outcomes

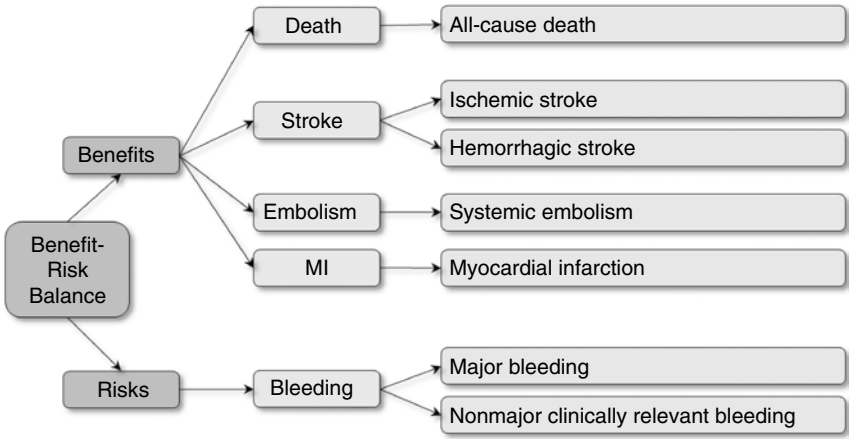
The second step of the BRAT B–R framework (see Table 35.1) is to identify and define the important outcomes: benefits, harms, and potentially other treatment properties of interest. While studies often measure many outcomes, many of these are highly correlated with each other or are causally dependent, some double-count events, and they may vary considerably in their clinical impact. Generally, B–R requires identifying a smaller set of key outcomes that drive the B–R assessment [11]. A tool often used to select and depict the outcomes used in B–R is a value tree, a hierarchic graph in which outcomes are grouped by anatomic, functional or clinical impact (Figure 35.1A). Identifying outcomes in the value tree also serves as a discussion tool in determining which outcomes are most important from patient and physician perspectives.

The value tree can also be utilized to understand problems caused by the interrelationships between outcomes. Outcomes that are easily interpreted individually can be problematic when considered collectively for B–R. For example, in the Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) study, the pivotal trial of dabigatran for atrial fibrillation, the primary efficacy outcome was stroke or systemic embolism, and the primary safety outcome was major bleeding [45]. Major bleeding is defined as the composite of fatal bleeding, critical organ bleeding (e.g., hemorrhagic stroke, intracerebral hemorrhage), transfusions ≥2 units packed red blood cells or whole blood, or hemoglobin drops ≥2 g/dL) [44]. Fatal and nonfatal hemorrhagic stroke events are included in both the primary efficacy and safety outcomes. This double-counting can lead to considerable confusion, as some events will count as both a benefit and a harm. Double-counting also occurs between all-cause death and other efficacy outcomes, as strokes, myocardial infarctions, and systemic emboli may be fatal. Counting these deaths twice can potentially distort the findings. Finally, major bleeding includes a mix of events that are fatal, cause irreversible harm or are transient without sequelae, yet each event is weighted equally in the composite outcome [44]. This large range of clinical impact under one outcome complicates the comparison between benefits and harms [46,47].

The value trees in Figure 35.1 show one way to resolve these issues [48,49,65]. Figure 35.1A shows a value tree using the key outcomes of a typical atrial fibrillation trial. Figure 35.1B's value tree includes the same events but with ischemic events classified only as benefits and hemorrhagic events classified only as risks. This provides a separation that mostly aligns with the treatment's mechanism of action and avoids any double-counting between benefits and harms (strokes with both ischemic and hemorrhagic components would need special consideration). Additionally, efficacy outcomes are defined to avoid double-counting. Finally, benefits and harms are classified by whether they are fatal or generally result in irreversible harm; fatal and critical organ bleeds are classified as fatal/irreversible, transfusions and hemoglobin drops are classified as reversible. This last classification is based on the FDA's approach to the approval of dabigatran and prasugrel, in which the FDA focused primarily on outcomes that were fatal or caused irreversible harm [50,51]. By separating the fatal and irreversible efficacy and safety events from less impactful ones, a first pass at the B–R assessment can be made with clinically comparable benefits and harms, then less impactful events can be included for additional detail in the assessment (see Figure 35.1B).

In some cases, having two or three value trees can be helpful. When there are different decision makers with different views on which events are most important, different value trees can be developed that reflect these different viewpoints, and the analyses conducted with each. For example, some decision makers may want to include

(A)



(B)

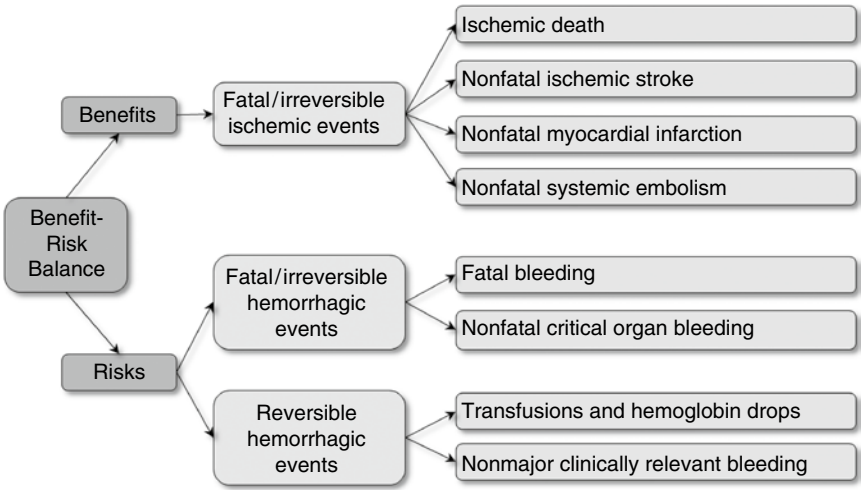


Figure 35.1 Example value trees for treatment of atrial fibrillation. (A) Endpoints from a typical atrial fibrillation statistical analysis plan. (B) Modified value tree with one approach for benefit–risk assessment. Major bleeding is defined as the composite of fatal bleeding, critical organ bleeding (e.g., hemorrhagic stroke, intracerebral hemorrhage), transfusions ≥2 units packed red blood cells or whole blood, or hemoglobin drops ≥2 g/dL) [44]. Nonmajor clinically relevant bleeding is defined as overt bleeding not meeting the definition of major bleeding but requiring medical intervention, contact with a health professional, a change in dosing of study drug/treatment or associated with discomfort or which impaired activities of daily living. MI, myocardial infarction.

some types of bleeding events besides critical organ bleeds under the irreversible harm category. In the Advanced Methods on the Horizon section, we outline a few approaches that sidestep some of these concerns of double-counting and multiple perspectives on which events are most important.

Identify and Summarize Source Data

The third step of the B–R framework differs considerably in preapproval (development) and postapproval contexts. In the preapproval phase, the choice of datasets for B–R is self-evident. The pivotal clinical trials must be used, and while the generally *post hoc* nature of B–R provides some flexibility, the protocol and statistical analysis plan prespecifies the populations and means for pooling data from multiple studies. Post approval, however, identifying and summarizing data for B–R has many more options and complexities.

Approaches for handling the choice of database, study design, and analysis approach that apply to pharmacoepidemiology in general also apply to B–R assessment. These topics are addressed in other chapters in this book, so we only briefly introduce important points regarding data sources more specific for B–R assessment, particularly as more data accumulate across the life cycle of a treatment.

The availability of electronic data sources, in particular claims data, registries, and electronic health records, has created new opportunities to conduct postapproval B–R assessments. This has also resulted in the increased discussion and utilization of pragmatic randomized trials and EHR-enabled prospective studies for the evaluation of effectiveness and safety, such as the Phase III pragmatic randomized Salford lung study in COPD [52,53]. With these electronic data sources, new challenges arise regarding evaluation of data quality and utility, and how to synthesize the data.

At the time of regulatory approval of a new medical treatment, B–R assessment is largely based on data from controlled clinical trials and preclinical studies. A necessary condition for the rigorous assessment of benefits and harms is the foresight to anticipate and accurately measure them. These studies often have a focus on well-characterized efficacy data. Specific safety outcomes may be collected based on anticipated harms due to the mechanism of action (for example, bleeding events in an anticoagulant trial). Otherwise, there is passive collection and reporting of adverse events by investigators if something appears to be abnormal.

With passive collection, it is therefore sometimes hard to distinguish between nonoccurrence of an event and underreporting of safety events due to informative censoring, early discontinuation, or inconsistent reporting. There are also harms that might not always be recorded in data sources, including "silent harms" that would be revealed with laboratory tests (for example, asymptomatic elevated liver enzymes that may be surrogates for future liver disease). Finally, analyses of serious, rare adverse events may not be adequately powered to detect differences between treatments. Thus, the absence of observed safety events or harms or their lack of statistical significance does not necessarily imply an absence of harm in a clinical trial or other setting.

There are analysis principles that can account for some of these limitations in the passive recording of safety events in clinical trials. For example, the intent-to-treat principle is used to capture important safety events resulting after early discontinuation or the end of the treatment period by following patients during these periods [54]. Sensitivity analyses of adverse event definitions may also be considered when it is suspected that a safety event may have been reported using different terminology. For example, community-acquired pneumonia in a trial in children might be recorded as community-acquired pneumonia, lower respiratory tract infection, and/or asthma exacerbation in the absence of a special adverse event collection

page, depending on the physician's judgment and accuracy of diagnosis. Further, the diagnosis of pneumonia might not be accurate, as there may be poor agreement on pneumonia diagnosis even when a chest X-ray is present [55–57]. In this example, the sensitivity and specificity of the pneumonia or lower respiratory infection event rates could be considered in the B–R assessment. However, additional studies may be required to investigate unexpected, rare but serious safety events noted in clinical trials to ensure adequate power and ascertainment, including subsequent trials or observational studies.

Following a medical treatment's approval, data informing B–R accumulate in a broader, larger sample of subjects in clinical practice and postapproval studies mainly through observational sources. B–R assessment is therefore often conducted by considering multiple studies conducted in a variety of settings, including randomized trials and observational studies. An understanding of strengths and weaknesses of study designs is important in weighing the B–R evidence across existing studies as well as in designing *de novo* observational or randomized studies.

We briefly describe important differences between randomized trials and observational studies that are important considerations for B–R assessment. Postapproval observational studies are more likely to compare two or more active treatments and observational studies generally do not contain a placebo arm, unlike many randomized controlled trials for registration. Some data limitations present in clinical trials also persist in the postapproval setting, including the passive reporting of safety data. Definitions of exposure and efficacy and safety outcomes in the observational settings are dependent upon the means of collection in the source data and generally vary from setting to setting. Medical diagnoses are typically recorded as a numeric medical code in the case of electronic sources as part of documenting healthcare encounters and/or processing medical insurance claims. Some diagnoses and symptoms might not be readily recorded in a claims or medical record, as they may not result in a healthcare visit (for example, stomach ache). Further, patient characteristics that may be important confounders or prognostic factors may not be available, particularly in healthcare claims data. For example, weight, smoking status, body mass index, and lung function tests may be available in an electronic medical record or patient registry but would not be reported in claims data as these data are not generally required to process healthcare claims. Efficacy or safety outcomes of interest may not be available or reported with the same granularity as required for a clinical trial. For example, specific types and severity levels of bleeding events that may be associated with an anticoagulant in observational sources are identified on the basis of medical codes [58] whereas more detail would typically be ascertained for bleeding events in randomized controlled trials [45]. More extensive discussion about validity of pharmacoepidemiologic data appears in Chapter 41.

The most robust data for B–R assessment are from randomized controlled trials and observational data sources where the important outcomes from the value tree have been rigorously captured and where confounding and selection bias have been rigorously accounted for (e.g., through randomization or statistical adjustment). A particularly important concern in observational data analysis is addressing confounding by indication or channeling bias, in which treatments are used differently depending on the degree of severity of the illness. For example, if a particular treatment is generally reserved for sicker patients, that treatment may appear to be associated with worse side effects than other treatments, simply because the sicker patients are more likely to have such side effects (see Chapter 43). Similarly, confounding by indication may mask the true benefit of a treatment

and even make the treatment look less efficacious than the comparator (see Chapter 33).

Merging the data quantitatively from many disparate sources and across different study designs is not always possible, desirable, or realistic. If comparisons are made between effect estimates across studies, the extent to which study populations, definitions, study designs, and analysis approaches are comparable should be described, including the potential for bias and confounding. Metaanalysis can be considered as a means to tabulate and combine effect estimates for benefits and harms when appropriate, using a transparent and systematic approach (see Chapter 36). Completed studies may sometimes be excluded from a metaanalysis on the basis of low methodological rigor. Metaanalyses can include direct or indirect comparisons between treatments depending on the availability of the data. A traditional metaanalysis approach would be applied when head-to-head studies have been conducted. For example, a metaanalysis of Phase III clinical trial data was conducted to compare oral anticoagulants to vitamin K antagonists in atrial fibrillation [59].

When head-to-head studies are not available, however, a network metaanalysis approach may be considered. For example, a network metaanalysis was conducted to compare oral anticoagulants relative to one another in atrial fibrillation [60]. The appropriateness of a metaanalytic summary might be questioned when results vary so much across studies that a single estimate loses interpretability. Methodological considerations for metaanalysis are further described in Chapter 36.

Customize the Framework

After identifying and summarizing source data, there may be a need to revisit the outcomes for B–R as described in Step 4 of the BRAT framework (see Table 35.1). Two common reasons for this revisiting are the emergence of new adverse events (AEs) not considered at the time the value tree was developed and differences between the B–R outcomes of interest and those available in the source data. New AEs may occur both during development and post approval, in which case additional harms can be added to the value tree. The available data do not always have the measurements desired or the granularity of outcomes to support the originally planned approach to B–R, particularly in claims and EHR data. In this case, the B–R assessment may require a revised set of outcomes. While constructing a formal value tree is simply a graphical exercise to facilitate selecting and communicating outcomes for B–R, customizing the outcomes to accommodate the limitations of the data sources may be necessary.

Consider the atrial fibrillation value tree in Figure 35.1B. The outcomes in this tree require distinctions such as fatal versus nonfatal myocardial infarctions and disabling versus transient major bleeds. If the available observational data sources cannot adequately measure these endpoints, the tree will need to be customized with proxy or less granular outcomes that are available. In some cases, an algorithm can be used to approximate a desired outcome. For example, primary hospital discharge diagnosis codes can be used to identify bleeding-related hospitalization and to differentiate upper and lower gastrointestinal bleeding [61]. The bleeding events in these algorithms may not fully align with the intended definition of major bleeding but may be sufficient, although different definitions make comparisons with RCT results potentially problematic.

Another approach used for atrial fibrillation is to simplify the value tree to two composite outcomes, potential thromboembolic events and intracranial hemorrhage, both defined by ICD-9-CM codes [62]. The risk differences for these outcomes can then be summed, potentially with a weight to reflect their relative importance, into what is typically called a "net clinical benefit" (NCB) outcome. There are

many variants of this approach, using different combinations of ischemic events and hemorrhagic events for the two composite outcomes, or some versions where both ischemic and hemorrhagic events are combined into a single composite endpoint [63–67]. An advantage of this approach is a more intuitive interpretation of the study outcomes compared to considering a large value tree. Using this approach, the weighting of two outcomes can be reduced to a single weight that is reflective of their relative clinical impact. This approach can also be implemented in nearly any clinical trial or observational dataset.

Disadvantages of using composite endpoints in this manner include that they are mixtures of disparate events that span the range of fatal or permanently disabling to transient with no sequelae. Comparing them without considering their components may fail to identify potentially important specific differences between treatments. Even with weights or patient preferences as described below, the comparison between outcomes with such a broad range of clinical impact can be difficult.

Assess Importance of Outcomes - Value Judgments and Patient Preferences

Step 5 of the BRAT framework (see Table 35.1) is to assess the relative importance of the outcomes. Consider an anticoagulant that reduces the chance of a cardiovascular death from 3% to 2% but increases the chance of a disabling hemorrhagic stroke from 1% to 5%. Is this B–R trade-off acceptable? The decision on whether the mortality benefit outweighs the hemorrhagic stroke depends not only on the changes in their chance of occurrence, but also on the importance one assigns to each event. These types of value judgments are a critical component of B–R.

Benefit–risk assessment is a combination of data-based probability assessment and value judgments [2–4,7,9,19,21,23]. Statistical analyses can provide the probability of different types of events prevented and caused, but they do not indicate how important those events are to decision makers [31]. In B–R, these value judgments are typically referred to as "weights" or "preferences." The preferences are most often for the different benefits and harms and other treatment characteristics, though they may also be assessed for treatments as a whole [68]. The measurement and application of preference is a growing field, with applications far beyond medical treatment B–R.

In many cases, these value judgments are based on clinical judgment. FDA and EMA B–R assessments reflect this approach [2,9]. For example, the FDA's anticoagulant B–R assessments described earlier use clinical judgment to partition events into two categories: events that are fatal or cause irreversible harm, and events that cause reversible harm [50,51]. This partitioning gives two effective "weights": clinically very impactful (very high weight) events and clinically much less impactful (much lower weight) events, and the B–R assessment is focused primarily on the very impactful events. Similarly, the EMA's guidance on review of drug applications (the rapporteur's day 80 critical assessment report) stresses the need to describe the value judgments used and notes that "a 'descriptive' approach with explicit considerations about the importance of the different effects and how trade-offs are weighed will generally be appropriate" [9].

Clinical judgment can be used to (i) rank endpoints in order of decreasing clinical impact, (ii) group them into categories such as the Common Terminology Criteria for Adverse Events (CTCAE) scale [69] or (iii) give rough decisions of what trade-offs are acceptable. These types of approaches are common not just in regulatory B–R, but in clinical practice. For example, physicians make similar judgment calls in determining which treatment might be best for an individual patient. Experts write treatment guidelines based on available scientific evidence,

clinical judgment, and, increasingly, patient preferences [70,71]. In a shared decision-making paradigm, physicians and patients discuss their B–R preferences for treatment, which may demonstrate that physicians and patients do not place the same weight on different outcomes [72,73].

While most B–R problems can be assessed with clinical judgment, there are many that are well served by formal, rigorous studies to assess how patients weigh benefits and harms [3,15,19,31,32,41]. Called "patient preference studies," these measure what attributes of a treatment are important to patients, the relative importance of these attributes, what trade-offs between them patients would accept, and the heterogeneity of these results among patients. Patient preference studies are taking on an increasing role in B–R and the literature in this discipline is growing [74–76]. Patient groups and pharmaceutical or medical device companies use them to incorporate the patient perspective into B–R. The FDA and the EMA are starting to consider preferences in decision making as complementary material along with efficacy and safety data. The Center for Devices and Radiological Health (CDRH) at the FDA has published a guidance document on the inclusion of patient preference information [4]. It has also conducted a patient preference study in surgical weight loss devices and approved a device in part based on results showing that a substantial proportion of the target population was willing to accept the risks associated with the device in exchange for the weight loss benefit [3]. In the FDA's recent approval of rituximab for the treatment of lymphoma and chronic leukemia, patient preference information assessed within the trial was included in a new Patient Experience section in the label [68]. The EMA has also piloted a preference study in patients with multiple myeloma [77], and both the FDA and EMA are conducting additional preference studies.

There are many different methodological approaches to elicit patient preferences, ranging from qualitative interviews through quantitative techniques, though in practice, a much smaller number is used in B–R assessment [16,19,29,30]. One of the most common quantitative methods is conjoint analysis, in which patients choose between sets of hypothetical treatments that have different combinations of benefit and harm attributes, and a regression analysis of the choices determines the relative importance of the attributes. Additional overviews of patient preference methods have been published [19,31,33,78].

There have been numerous preference studies for anticoagulants, in both patients and other stakeholders [75]. In general, atrial fibrillation patients are willing to accept certain bleeding risks for a decrease in the probability of experiencing a stroke. However, there is substantial variability in the threshold number of bleeds observed for the acceptance of an oral anticoagulant. Figure 35.2 shows the results of a preference survey for acute coronary syndrome [72]. The bars in the figure show the relative importance US patients place on death, different severities of stroke, myocardial infarction, different degrees of bleeding, and angina. US patients regarded nonfatal disabling stroke as equal in importance to death. The low height of the bars on the right indicates that patients are willing to accept considerable probability of bleeding, angina or even nonfatal heart attack in exchange for treatments that reduce the chance of death or disabling stroke. In a different preference study for atrial fibrillation, in exchange for reducing the chance of nonfatal disabling stroke by one percentage point, US patients would accept up to a 6.3% chance of nonmajor clinically relevant bleeding and up to a 2.0% chance of extracranial major bleeding [79].

In addition to population averages of preference as show in Figure 35.2, preference studies can measure heterogeneity in preferences. Some patients will be very tolerant of risks

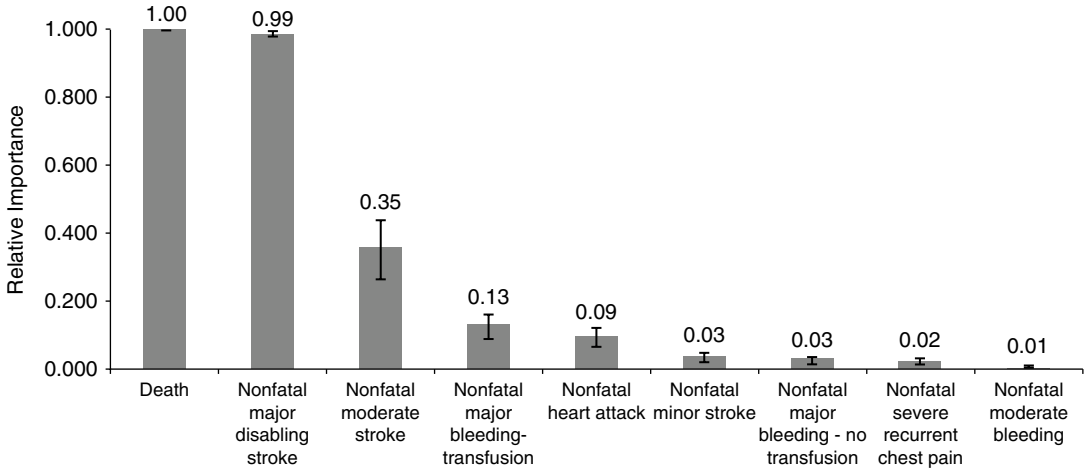


Figure 35.2 Relative importance to US patients of key treatment-related benefits and harms for acute coronary syndrome. The terminology for the benefit and harm attributes reflects the patient-friendly language used in the survey; for example, recurrent chest pain refers to angina. Bars show the relative importance of each attribute, with the most important attribute scaled to 1.0. *Source:* Yuan *et al*. [72]. Reproduced with permission of Taylor & Francis.

while others will be very risk averse. Preference studies can measure how widely patients' preferences vary and whether there are distinct subgroups of patients with important differences in preferences. While often referred to as patient preferences, the preferences of physicians, caregivers, and others can be ascertained and compared with these methods. For example, a preference study of patients and physicians was conducted in the US and Japan on the relative importance of benefits and harms of anticoagulant therapy using a conjoint analysis experiment. Japanese physicians and patients exhibited different preferences from each other, with physicians being less likely to tolerate the risk of bleeding in exchange for the reduction in the risk of stroke [79]. These results may in part explain lower prescribing rates of anticoagulants for atrial fibrillation patients in Japan compared with the US [80].

As popular as patient preference studies are becoming for B–R applications, there are many unanswered questions in terms of requirements in the design, conduct, and analysis of these studies for regulatory and health technology assessment purposes. We describe ongoing initiatives to address these questions at the end of the chapter.

Putting It All Together: Integrated B–R Assessment

Because synthesizing the evidence in a structured manner is often tightly linked with visualizing the data or analyses, we discuss integrated B–R assessment and communicating the assessment, Steps 6 and 7 of the BRAT B–R framework (see Table 35.1), together.

There are numerous approaches to combining data on medical need, outcome data, uncertainty, clinical judgment, preferences, heterogeneity, etc. into a B–R assessment [18,21,23,81–85]. Benefit–risk approaches are often categorized as qualitative, semiquantitative or quantitative. Qualitative approaches use textual descriptive summaries to make a B–R argument. They are most applicable when the assessment is self-evident from the data, such as statistically significant benefit and no appreciable AEs. Semiquantitative approaches use a combination of

tabular and graphical displays, potentially coupled with preference or weight information. Most B–R in regulatory and practical applications is descriptive or semiquantitative, and we focus on these approaches here [2,7,9]. Quantitative approaches compute summary metrics by combining the data and potentially preferences from multiple outcomes. We touch upon quantitative methods here but address them more fully in the Future section.

An increasingly common means to display B–R data is an effects table [9,81] which summarizes information on all key benefits and harms (Table 35.2). In most versions, there is one row for each benefit or harm. The columns vary but may include the outcome name, a brief outcome definition and units, the outcome value for each treatment, estimates of between treatment differences (e.g., risk difference) with associated uncertainty (e.g., 95% confidence intervals), brief notes on strength of evidence and a link to data sources. While relative measures such as relative risk, odds ratio or hazards ratio can also be included, these measures are generally less useful for B–R. Because the baseline rates for outcomes may be disparate, a large relative risk may correspond to a very small absolute difference in the number of events when the baseline rate is low, while small relative risks may correspond to large absolute differences when the baseline rate is high.

Table 35.2 shows an effects table for the atrial fibrillation value tree in Figure 35.1B. The data are simulated but realistic, and the term "study drug" is a proxy for any medical intervention. All outcomes in the example presented are in person-year rates, though effects tables in general can include any type of risk or rate calculation and can

Table 35.2 Effects table for atrial fibrillation (simulated data). All outcomes are measured per 10 000 person-years.

Outcome	Event rate (/10 000 person-years)		Rate difference /10 000 person-years (95% CI)	NNT or NNH
	Study drug	Comparator		
Efficacy				
Ischemic death, ischemic stroke, MI or systemic embolism	430	512	−87 (−143, −30)	−115
Ischemic death	189	221	−34 (−72, 3)	−294
Nonfatal ischemic stroke	154	172	−19 (−51, 13)	−526
Nonfatal myocardial infarction	83	100	−17 (−43,8)	−588
Nonfatal systemic embolism	4	20	−16 (−26, −7)	−625
Safety				
Major bleeding	398	345	49 (0, 99)	204
Fatal and critical organ bleeding	119	125	−6 (−33, 20)	−1667
Fatal bleeding	37	38	−1 (−15, 13)	−10000
Nonfatal critical site bleeding	82	87	−5 (−31, 21)	−2000
Transfusions > =2 units or Hbg drop > − 2 g/dL	308	245	60 (17,103)	167
Clinically relevant nonmajor bleeding	1192	1137	44 (−50. 138)	227

CI, confidence interval; MI, myocardial infarction; NNT, number needed to treat; NNH, number needed to harm. Positive NNTs or NNHs indicate more events on the study drug. Negative NNTs or NNHs indicate more events on the comparator.

show categorical or continuous outcomes. Because the incidence rate of most outcomes is low, the data are scaled to a hypothetical population (10 000 person-years) to simplify representation and comprehension of these data [86]. The rate differences represent the additional number of events caused or prevented by using one treatment compared to another treatment after 10 000 person-years exposure. With many outcomes considered in B–R, we generally show the 95% confidence interval (CI) as a measure of uncertainty but not for statistical hypothesis testing, with the exception of those endpoints for which a study was powered. Number need to treat (NNT) and number need to harm (NNH) are also included in the table. NNT is the number of person-years exposure with a treatment versus comparator to prevent one additional harmful event. NNH is the number of person-years exposure with a treatment versus comparator to cause one additional harmful event. NNT and NNH are calculated as the reciprocal of the corresponding rate differences [87,88].

Focusing first on events that are fatal or cause irreversible harm, per 10 000 person-years, there are 87 (95% CI 30, 143) fewer events per 10 000 person-years in the efficacy composite of ischemic death, ischemic stroke, MI or systemic embolism, and there are six (95% CI -20, 33) fewer fatal and critical organ bleeding events on the study drug. Note that the sign of the rate difference and confidence interval of the safety outcome is reversed when referring to fewer events on the study drug. The composite efficacy outcome is statistically significant favoring the study drug, while the fatal and critical organ bleeding shows no meaningful difference, suggesting benefits outweigh harms.

When considering the full set of outcomes, a forest plot of the rate differences in Table 35.2 is very helpful, particularly when communicating a B–R assessment (Figure 35.3). At a glance, the plot makes clear that death, ischemic stroke, myocardial infarction, and non-CNS systemic embolism each contribute meaningfully to efficacy. Major bleeding favors the comparator, but the forest plot makes it clear that this difference is driven primarily by less impactful bleeds. There is no difference in fatal and critical organ bleeding, while the comparator is superior in the less impactful transfusions, hemoglobin reductions as well as nonmajor clinically relevant bleeding (see Figure 35.3). The trade-off is preventing 87 (95% CI 30, 143) fatal/irreversible harm ischemic events versus causing 60 (95% CI 17, 108) reversible bleeding events, both per 10 000 person-years. Without taking patient preferences into account, in this simulated example, the benefits appear to outweigh the harms.

This B–R analysis can also be performed with NNT and NNH. Comparing the broadest efficacy and safety outcomes, one harmful ischemic event is prevented for every 115 person-years of exposure to the study drug versus comparator (NNT), while 204 person-years exposure is needed to see an excess major bleeding event (NNH) on study drug. That is, less study drug exposure is needed to achieve a beneficial effect than a harmful effect. Additionally, each beneficial event is far more clinically impactful than each harmful event, since there were few fatal bleeds or critical organ bleeds. Taken together, these points strongly suggest benefit exceeds harm. Additionally, the NNH for fatal and critical organ bleeding is -1667. The large absolute value indicates that the between-treatment difference is small. The negative value indicates that the difference favors the study drug despite the bleeding being classified as a harm. In this case, these three outcomes are sufficient to assess B–R. However, when considering many outcomes at once, NNT and NNH are less helpful. The techniques for comparing many NNTs and NNHs simultaneously are mathematically complex and require weights or preferences for each outcome [84,89]. Additionally, confidence intervals for NNT and NNH can be difficult to interpret when

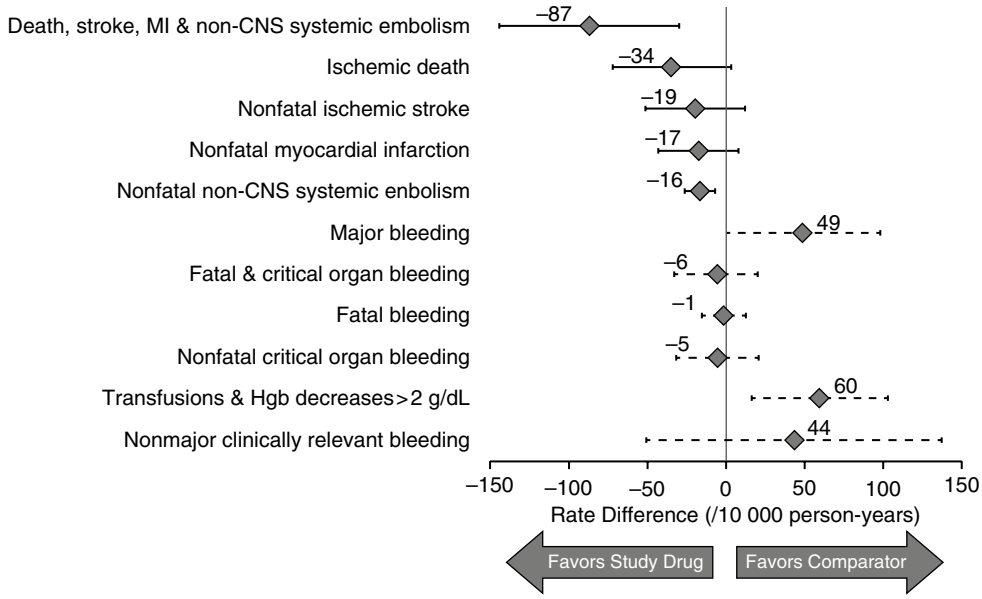


Figure 35.3 Forest plot showing rate differences per 10 000 person-years for key benefits and harms in atrial fibrillation treatment (simulated data). Diamonds are points estimates. Bars show 95% confidence intervals. Efficacy outcomes are shown as solid lines. Safety outcomes are shown as dashed lines. CNS, central nervous system; Hgb, hemoglobin; MI, myocardial infarction.

outcomes are not statistically significant [87,90,91]. For these reasons, NNT and NNH have limited application in B–R assessment.

Weighting and the Patient Perspective

The integrated B–R approaches reviewed above used clinical judgment to make a defensible B–R decision. The example in Table 35.2 and Figure 35.3 lends itself to such approaches. However, when the B–R trade-off is more complex, such as when some key outcomes favor the study drug and other key outcomes favor the comparator, weighting and preference assessments can be critical, both to incorporate the patient perspective and to make a B–R assessment.

There are numerous approaches by which the weighting or patient preferences can be incorporated into B–R. Figure 35.4 shows a forest plot with outcomes placed in order of decreasing clinical impact for patients, where clinical impact is based on standard gamble and time trade-off utilities (a type of weight) obtained from the Tufts Cost-Effectiveness Analysis Registry. In this approach, composite outcomes are not included when their components are displayed. The most severe outcomes all favor the study drug, while the least severe favor the comparator. Again, these data suggest that the benefits outweigh harms. An advantage of using the rank rather than actual weight or preference values for outcomes is that there is less opportunity for disagreement among decision makers about the values of the weights or their exact order. For example, in Figure 35.4, the five outcomes associated with the most severe clinical impact all favor the study drug. The B–R assessment is the same regardless of the ranking within the most severe outcomes.

Another class of approaches that incorporate weights or the patient perspective in B–R are

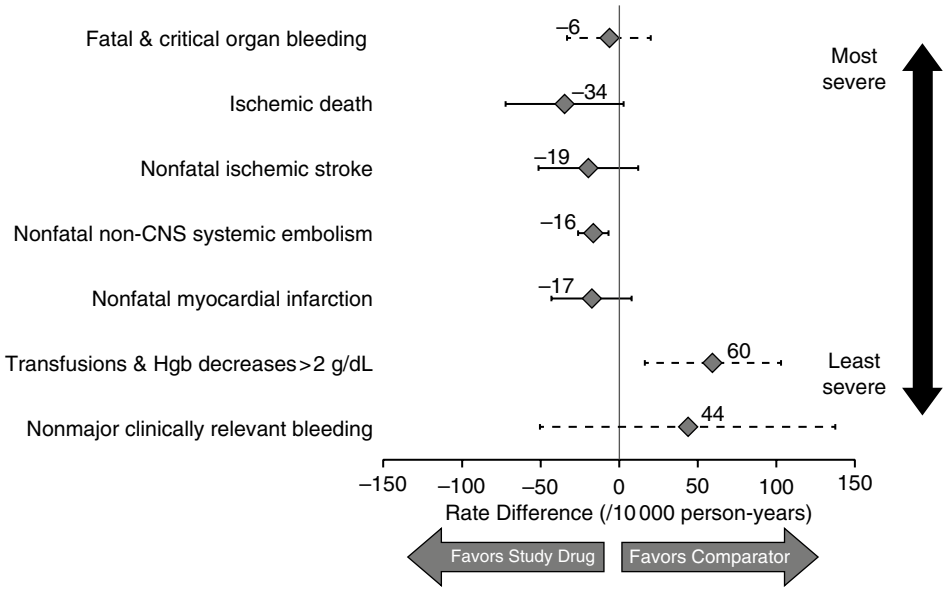


Figure 35.4 Forest plot for key benefits and harms in atrial fibrillation treatment ranked in order of decreasing clinical impact as measured by utility. Diamonds are points estimates. Bars show 95% confidence intervals. Efficacy outcomes are shown as solid lines. Safety outcomes are shown as dashed lines. CNS, central nervous system; Hbg, hemoglobin.

net clinical benefit (NCB) measures. While the terminology is not standard, we define NCB approaches as those that use a weighted sum of risk (or rate) differences between study drug and comparator to summarize the difference between treatments:

$$NCB = \sum_{i=1}^{N} w_i \times RD_i$$

where N is the number of outcomes, w_i is the weight associated with outcome *i*, and *RDi* is the between-treatment rate difference or risk difference for outcome *i*. Weights can be obtained from clinical judgment, clinical sequelae or any of the wide variety of patient or physician preference studies described above. Typically, beneficial events have positive weights and harmful events have negative weights. The larger the absolute value of the weight or the larger the risk difference for an outcome, the more that outcome contributes to NCB.

As a simple example, the NCB approach described above for atrial fibrillation uses two composite outcomes: potential thromboembolic events and intracranial hemorrhage [62]. In this case, net clinical benefit is defined as:

$$NCB = w_{TE} \times RD_{TE} + w_{ICH} \times RD_{ICH}$$

where RD_{TE} and w_{TE} are the risk difference and weights for thromboembolism, and RD_{ICH} and w_{ICH} are the risk difference and weight for intracranial hemorrhage. Risk difference here is defined as the probability of events on study drug minus that on comparator. Weights of 1.0 for potential thromboembolic events and 1.5 for intracranial hemorrhage were based on mortality and disability data from a large observational atrial fibrillation study [92], indicating that the clinical impact of a typical ICH is weighted 1.5 times that of a typical thromboembolic event. Since $w_{TE} = 1$, a unit change in NCB is equivalent to a unit change in the probability of thromboembolism. Hence,

this net clinical benefit is also in units of thromboembolic events. For example, if $RD_{TE} = -80$ (95% CI −85, −75) per 10000 patient-years and $RD_{ICH} = 30$ (95% CI 25, 35) per 10000 patient-years, the study drug is better than the comparator on thromboembolic events and worse on ICH events. Net clinical benefit is $-80 + 1.5 \times 30 = -35$ (95% CI −49, −21) thromboembolic event equivalents per 10000 patient-years. Thus, the difference between treatments is the equivalent of 35 fewer thromboembolic events per 10000 patient-years, favoring the study drug.

There are many other ways in which a net clinical benefit measure can be defined. For example, an NCB measure can be based on a weighted sum of the seven outcomes' rate differences from Figure 35.4, with weights obtained from the utility scores used to sort the outcomes in the figure or from a preference study such as in Figure 35.2. Sensitivity analyses can be conducted on NCB measures, where distributions for the weights are propagated into uncertainty in the NCB result. The distribution of NCB, whether the uncertainty is based on clinical data uncertainty, weight uncertainty or both, can give more complex metrics such as the probability that benefit exceeds risk (example in appendix 3 of [93]).

A more general approach to B–R assessment is multicriteria decision analysis (MCDA) [40–42,94]. MCDA is a general and flexible quantitative approach to decision making that can accommodate not just the dichotomous outcomes described earlier for NCB but any type of input, including continuous outcomes, categorical outcomes, and properties that may need specialized scales such as ease of use and drug–drug interactions. This flexibility comes at the cost of greater complexity, and MCDAs are generally reserved for more complex B–R problems.

Multicriteria decision analysis models have been applied to many B–R problems, including atrial fibrillation [95–98]. The EMA has started considering their use for complex B–R problems [41,77]. Figure 35.5 shows the results of an MCDA model for natalizumab, a treatment for multiple sclerosis [97]. Despite the very high weight (left column in Figure 35.5) given the treatment-related adverse event of progressive multifocal leukoencephalopathy (PML), the very small difference in between-treatment rates (middle column) causes PML to have negligible impact on the B–R analysis (right column). In contrast, despite its much smaller weight, relapse's large between-treatment difference causes it to be the main driver of the B–R balance. In this manner, MCDA models can help show the relative contributions of rates and weights for many outcomes. As for NCB approaches, MCDA models can also be extended with distributions for both weights on clinical inputs, often called stochastic multicriteria acceptability analysis, providing probabilistic assessments and sensitivity analyses for B–R results [99,100].

Preference studies can also be used to address the questions raised at the outset of this chapter – what probability of harm is acceptable in exchange for a given degree of benefit? A preference study can provide the maximum acceptable risk, that is, the maximum probability or level of severity of a harm that a patient will accept in exchange for a given benefit. Similarly, preference studies can provide the minimum acceptable benefit, that is, the minimum benefit that a patient will require in exchange for a given probability or severity of a harm. For example, on average, patients with mild cognitive impairment are willing to accept up to about a 5% chance of death or disabling stroke in exchange for reducing moderate dementia symptoms to mild symptoms for one year [101]. In atrial fibrillation, in exchange for reducing the chance of nonfatal disabling stroke by one percentage point, US patients on average would accept over a 6.0% chance of nonmajor clinically relevant bleeding and up to about a 2.0% chance of extracranial major bleeding [79].

Patient preferences are heterogeneous – different patients would accept different tradeoffs in exchange for the same benefit [4].

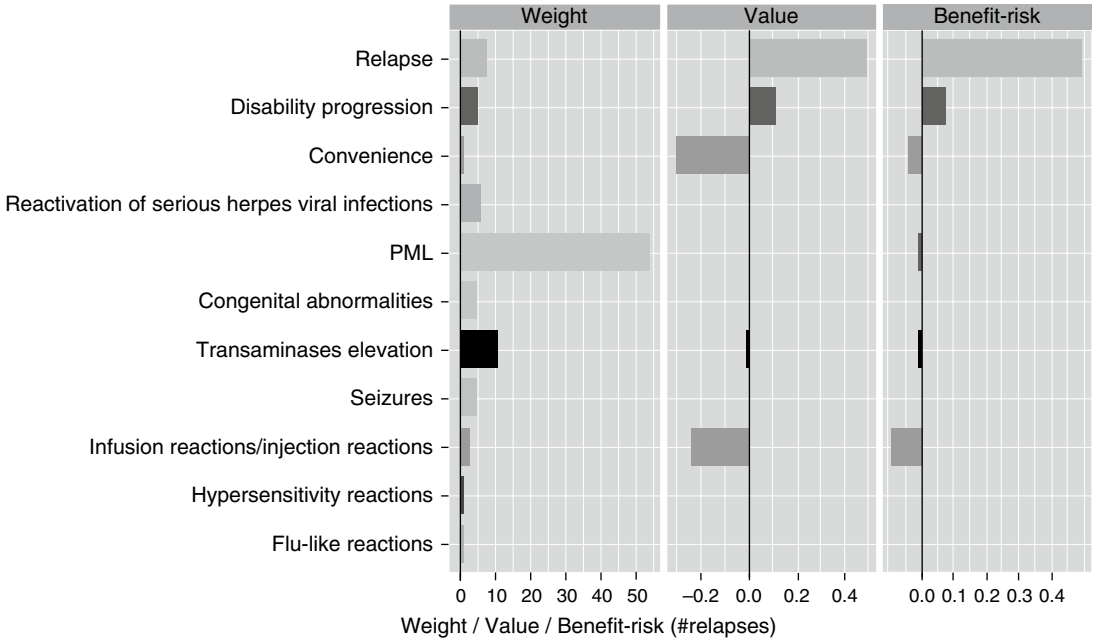


Figure 35.5 Output of multicriteria decision analysis (MCDA) model for natalizumab versus placebo. Weights (*left column*) reflect the relative importance of each outcome, value (*middle column*) reflects the risk difference between natalizumab and placebo on a normalized scale, and benefit–risk (*right column*) is the product of weight and value, reflecting the contribution of each outcome to the overall B–R assessment [97]. PML, progressive multifocal leukoencephalopathy.

Preference studies can measure this heterogeneity and identify subgroups of patients with different preferences. The balance of benefits and harms will differ in these subgroups, and preference studies can be used to assess the proportion of patients that would choose to use a given treatment – an assessment that was in part the basis for approval of a medical device to treat obesity [3]. Finally, preference studies can be used to understand the different perspective on acceptable B–R trade-offs in different stakeholder groups. For example, a preference study in patients with schizophrenia and psychiatrists who treat such patients showed that patients regarded improving the positive symptoms of schizophrenia as the most important benefit to a far greater degree than did their psychiatrists [102].

This section has covered the more common approaches to integrated B–R assessment. There are many ongoing initiatives advancing these methodologies and developing policy for their use in regulatory and other applications. An important role for these initiatives is addressing limitations in these methods. For example, there is a growing set of approaches for B–R that account for dependency between benefits and harms, consider competing risks, and combine patient-reported outcomes and patient preference studies. There are also numerous other approaches to communicating B–R [2,81,103,104]. In the final section, we address the ongoing initiatives to improve methodology and policy as well as some of the newer approaches for B–R.

The Future

Maturation of the Field: Initiatives, Guidances, and Partnerships

The many advances in structured approaches to B–R assessment and patient preference studies have also opened many lines of inquiry on methodology and policy. There are numerous ongoing research groups and public–private partnerships seeking to address these open questions.

As described at the start of the chapter, both the FDA and EMA have implemented a structured approach to communicating their B–R assessment. By 2020, the FDA is expected to publish guidance on B–R assessment for new drugs and biologics, focusing on its B–R framework [5,28]. Additional FDA guidance will focus on systematic approaches to collecting and utilizing patient and caregiver input for regulatory decision making, including patient preference studies [5,105,106]. Along with the updated approach to B–R in the ICH Clinical Overview template, this guidance will provide a foundation for FDA expectations in B–R assessment and reporting.

Public–private partnerships have an important role in aligning industry, regulatory agencies, patient groups, and academia on methodology and policy. The Innovative Medicines Initiative (IMI), Europe's largest public–private partnership, conducted IMI PROTECT, a large project that developed recommendations for B–R methodology and communication [16,81,97,103]. The Medical Device Innovation Consortium (MDIC) developed a framework on the assessment and use of patient preference studies in medical device development and regulatory review [19]. This framework is one of the key background documents supporting the FDA's guidance on patient preference studies [4]. The IMI is currently funding the IMI PREFER project that continues the MDIC work and will generate expert and evidence-based recommendations for industry, regulators, and health technology assessment groups on the assessment and use of patient preference studies for decisions in the medical treatment life cycle [15]. The Quantitative Sciences in the Pharmaceutical Industry (QSPI) Benefit–Risk Working Group has developed several novel statistical approaches to B–R [107]. These are just some of the key recent or ongoing initiatives.

Academic organizations such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), the Society for Medical Decision Making (SMDM), and the International Academy of Health Preference Research (IAHPR) continue to develop standards for B–R decision making and the proper conduct and use of patient preference studies.

Finally, more and more patient advocacy groups are taking on a role in B–R. For example, Parent Project Muscular Dystrophy collaborated with a scientist from Johns Hopkins University to conduct preference studies in muscular dystrophy and developed a monograph on B–R in rare diseases [108]. Other patient groups in Parkinson disease, diabetes, lung cancer or other disease areas are doing similar work. These groups are potential collaborators for academic, industry or regulatory patient preference and B–R work.

Advanced Methods on the Horizon

Benefit–risk assessment is an evolving field that is critical to optimal medical and public health decision making. In the future, there will be increasing demand for and reporting of B–R assessment. Assessment will become increasingly more structured, systematic, and quantitative as methodologies improve and become widely accepted. Areas of emphasis will include (i) the refinement of B–R assessment tools and application techniques, (ii) development of methodologies with greater pragmatism and applicability to clinical practice, and (iii) transitioning B–R assessment from a *post hoc* exercise to one that has a prespecified methodology or, potentially, a testable hypothesis.

The evolution of improved means to communicate B–R must continue to progress. Greater emphasis will be placed on risk difference summaries rather than relative risks or hazard ratios for greater interpretability. For example, risk differences as measured by differences in Kaplan–Meier estimates are more appropriate B–R assessment summaries than hazard ratios obtained from proportional hazards regression [109,110]. Graphical summary displays will be used with greater frequency to more effectively communicate complex and multidimensional effects of treatments [103,104]. And B–R frameworks will provide the succinct textual summary that accompanies these displays.

Despite being the gold standard for evaluating the safety and effectiveness of treatments, clinical trials may have shortcomings in informing medical decision making [111]. A key goal of B–R assessment is to contrast the benefits and harms of therapeutic alternatives as they are experienced by patients to help inform medical decision making. However, typical summaries of the study data for B–R assessment treat benefits and harms independently, then compare outcomes using either clinical judgment or integrating them using NCB or MCDA types of approaches. B–R assessment based on combining the separate marginal effects of each outcome in this fashion has limitations, as it fails to (i) account for any associations between outcomes (e.g., patients who benefit from the treatment may not experience any treatment-related harms, while those who do not benefit may experience the harms), (ii) systematically incorporate the relative importance of combinations of outcomes (e.g., the weight associated with a rash and pain experienced simultaneously may not be the same as the sum of the weights for rash and pain independently), or (iii) effectively deal with competing risks (e.g., trials that censor patients once they experience certain adverse events may not track subsequent benefits or harms that patient may experience).

As an example, suppose 100 patients are treated with a new treatment versus placebo, and one efficacy outcome and one safety outcome are measured and considered of equal importance (weight). Further suppose the efficacy and adverse event rates for the new treatment are both 50%, while the rates for both are zero for placebo. If the 50 patients who experience a treatment benefit (i.e., the efficacy event) are the same patients who experienced harm (i.e., the safety event), then the net clinical benefit is zero for all patients and the drug is worthless (Figure 35.6A). However, if the 50 patients with benefit are different from the patients who experienced the harm, then the net clinical benefit is positive for half the patients and negative for half the patients; that is, it is a good drug for half of the patients and a bad drug for the other half (Figure 35.6B), with the accompanying challenge of being able to predict which individuals fall into each category. Marginal summaries on the benefit and harm separately cannot distinguish these two scenarios, emphasizing the information

(A)

	Benefit	No Benefit
Harm	50%	…
No Harm	…	50%

(B)

	Benefit	No Benefit
Harm	…	50%
No Harm	50%	…

Figure 35.6 Demonstration of the impact of dependencies between outcomes. Both tables show risk differences for one benefit and one risk. The benefit and harm are assumed to be equally weighted. Part A shows the case where the treatment benefits no one (NCB=0 for all patients); Part B shows the case where the treatment benefits half of patients (NCB >0 for 50 patients and NCB <0 for 50 patients). However, the marginal distributions of benefit and harm of the two tables are identical.

lost by considering the outcomes independently. Furthermore, in most clinical trials, efficacy and safety are typically evaluated in different analysis populations, challenging the ability to generalize results when aggregating them. New methodologies for B–R assessment that have greater practicality in translating to clinical practice are needed. There are several on the horizon.

Earlier approaches to account for benefits and harms within a patient included a graphical representation of B–R over time. The status of a patient at any time during the trial is depicted using one of five health states: benefit without AE, benefit with AE, neither benefit nor AE, AE with no benefit, and withdrawal [112]. Using a different color for each health state, each patient's B–R pattern is represented using colored bars. By stacking all patient B–R data over time, sorting by different health states and comparing the pattern for a treatment and comparator, the B–R for the sample overall could be ascertained by visual inspection. This approach allows for dependency between two outcomes – sufficient for the example in Figure 35.6 but not for much more complex problems. As acknowledged by the authors, this graphical representation of B–R over time offers a relatively limited view of the data but it can contribute to B–R assessment. A Bayesian approach to longitudinal B–R assessment based on this representation allows for several net clinical benefit types of B–R endpoints [113]. The primary advantage of the Bayesian approach for longitudinal data is its ability to borrow information from prior assessments and recursively update posterior estimates of B–R measures. This approach can also be extended to any number of endpoints. Other Bayesian approaches to B–R are reviewed in Costa *et al.* [114].

The Desirability of Outcome Ranking (DOOR) is an evolving novel approach to B–R assessment that also accounts for benefits and harms within an individual patient. It is pragmatic from the perspective of informing medical decision making [115] and addresses the limitations described above. Conceptually, DOOR uses outcomes to analyze the patients rather than the patients to analyze the outcomes. A key to utilizing DOOR is to determine how to analyze one patient before analyzing many. For example, consider the cardiovascular example discussed earlier where outcomes included mortality, stroke (disabling and nondisabling), MI, and bleeding events. Traditional B–R assessment in cardiovascular trials often use time-to-first-event analyses without incorporating the cumulative nature of multiple events, the association between events, the competing risk of death or the weight of multiple events in a patient. DOOR can address these issues. Consider the experience of each study participant in terms of the events that they may experience. Suppose that mortality is considered the worst outcome and disabling stroke is considered the next worst outcome. Nondisabling strokes, MIs, and bleeds are harmful events but are generally somewhat transient and thus may have similar importance to each other but less than that of a disabling stroke. More of any of these events is worse than fewer events. Zero events is the best-case scenario. Thus, DOOR for this setting may classify global patient outcomes into one of, for example. five mutually exclusive categories from most desirable to least (Table 35.3). The distribution of DOOR between competing treatment options can be compared.

Two distinct analyses of DOOR can be conducted. One approach consists of conducting all possible pairwise comparisons of patients between treatments. One can then estimate the DOOR probability, the probability that a randomly selected patient randomized to one treatment has a better DOOR than a randomly selected patient on the other treatment. This may have an intuitive appeal from a clinical perspective given the connection to a common and important question that arises during clinical practice decision making: what is the probability that a patient will have a better overall

Table 35.3 Example desirability of outcome rankings (DOORs) for an anticoagulant study.

Rank	Outcome
1	Survived with no events
2	Survived with one nondisabling stroke, myocardial infarction (MI) or bleed
3	Survived with more than one nondisabling stroke, MI or bleed
4	Survived with disabling stroke
5	Death

response on one treatment relative to another? The win ratio, the relative frequency with which a randomly selected patient randomized to one treatment has a more desirable versus less desirable DOOR compared to the other treatment, can also be estimated. In the second type of analysis, "partial credit" levels of the ordinal outcome are scored similar to an academic test [116]. The most desirable outcome (survived with no events) is scored as 100 and the least desirable (death) is scored as 0. Middle categories are given partial credit. The amount of partial credit may be obtained through patient preference studies or a survey of expert clinicians. Sensitivity analyses to varying partial credit scoring permits personalized analyses. These analyses were used to compare colistin with ceftazidime-avibactam in the treatment of infections due to carbapenem-resistant enterobacteriaceae [117].

The bond between B–R assessment and precision medicine will also continue to grow (see Chapter 30) and better connect with shared decision making. After all, the goal in practice is not to identify who will benefit or who will be harmed but who has a positive B–R profile. Benefit–risk assessment should include sensitivity analyses to investigate and illustrate how the B–R profile varies across patient subgroups defined by demographics, genetics or baseline disease status, and a variant of DOOR called DOOR subgroup treatment effect pattern plot (DOOR STEPP) displays the distribution of DOOR as a function of baseline patient characteristics [118].

Incorporation of patient preferences and values will be necessary for increasingly quantitative and systematic evaluation. Preference studies tied to specific trials and disease areas will become more prevalent. Sensitivity analyses to patient preferences should become far more common, assessing the robustness of the results for regulatory assessment and to allow for personalized assessment and decision making in practice.

As the science and practice improve, B–R assessment may begin transitioning from a *post hoc* exploratory nonsystematic evaluation to one that is prespecified, systematic, and a primary focus. For example, some studies have incorporated prespecified systematic formal B–R outcomes and analyses into clinical trial protocols. This can be challenging in trials developing novel treatments due to the potential emergence of previously unknown adverse events. However, this has been addressed in some instances by focusing on how patients feel, function, and survive in a general sense rather than a focused evaluation of specific adverse events. Patient preference substudies and the collection of quality-of-life information will be included more frequently in clinical trials rather than being separate and distinct evaluations. When formal B–R hypotheses cannot be prespecified, the planned methodology for assessing B–R can be prespecified, such as the use of risk differences rather than hazard ratios, the use of outcomes that avoid double-counting, and the use of an NCB or DOOR approach once the key benefits and harms are known after unblinding study results.

Benefit–risk assessment is also critical for the evaluation and comparison of diagnostic tests. Different diagnostic errors carry different

clinical consequences which should be carefully evaluated and incorporated into B–R evaluation of diagnostic application. To address this, Benefit-Risk Evaluation of Diagnostics: A Framework (BED-FRAME) was proposed as a strategy for pragmatic B–R evaluation of diagnostics using weighted accuracy and plots of diagnostic yield [82,119]. The average weighted accuracy (AWA), a composite B–R summary measure, is being used as the primary measure in a clinical performance study evaluating the utility of a host response-based diagnostic test categorizing acute respiratory tract illness into bacterial, viral, or neither etiology [120].

Conclusion

Assessing the B–R balance of medical treatments has always been an integral part of drug development, regulatory, and public health decisions. Methodology for B–R has advanced considerably, and health authorities worldwide have developed much more rigorous B–R regulatory policies relating to regulatory approval and post-approval decisions. These advances have led to transparent and rigorous systematic approaches to B–R which incorporate the patient perspective and can be communicated clearly and succinctly to clinical and patient audiences.

References

1 Understanding the Benefits and Risks of Pharmaceuticals: Workshop Summary. Washington DC: Institute of Medicine, National Academies Press, 2007.
2 FDA. Structured Approach to Benefit–Risk Assessment in Drug Regulatory Decision-Making. Draft PDUFA V Implementation Plan 2013. www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf (accessed April 25, 2019).
3 Ho MP, Gonzalez J, Lerner H, *et al*. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc* 2015; **29**(10): 2984–93.
4 FDA CDRH. Guidance for Industry and Food and Drug Administration Staff: Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling 2016. Silver Spring: FDA.
5 FDA. PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2018 Through 2022. www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM511438.pdf (accessed April 25, 2019).
6 FDA. Enhancing Benefit–Risk Assessment in Regulatory Decision-Making. www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm326192.htm (accessed April 25, 2019).
7 FDA, CDRH. Guidance for Industry and Food and Drug Administration Staff: Factors to Consider When Making Benefit–Risk Determinations in Medical Device Premarket Approval and De Novo Classifications. Silver Spring: FDA.
8 EMA. Benefit–Risk Methodology Project. www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/document_listing/document_listing_000314.jsp#section1.
9 EMA. Guidance document on the content of the <Co-> Rapporteur day <60><80> critical assessment report. www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2016/05/WC500206989.pdf (accessed April 25, 2019).
10 International Conference on Harmonisation. Periodic Benefit–Risk Evaluation Report (PBRER) E2C(R2). www.ich.org/

fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2C/E2C_R2_Step4.pdf (accessed April 25, 2019).
11 Revision of M4E guideline on enhancing the format and structure of benefit–risk information in ICH, Efficacy – M4E(R2). www.ich.org/fileadmin/Public_Web_Site/ICH_Products/CTD/M4E_R2_Efficacy/M4E_R2__Step_4.pdf (accessed April 25,2019).
12 Noel R, Hermann R, Levitan B Application of the Benefit-Risk Action Team (BRAT) framework in pharmaceutical R&D: results from a pilot program. *Drug Inform J* 2012; **46**(6): 736–43.
13 Levitan BS, Andrews E, Gilsenan A, *et al*. Application of the BRAT framework to case studies: observations and insights. *Clin Pharmacol Ther* 2011; **89**(2): 217–24.
14 Coplan PM, Noel R, Levitan B, *et al*. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. *Clin Pharmacol Ther* 2011; **89**(2): 312–15.
15 de Bekker-Grob EW, Berlin C, Levitan B, *et al*. Giving patients' preferences a voice in medical treatment life cycle: the PREFER Public–Private Project. *Patient* 2017; **10**(3): 263–6.
16 Mt-Isa S, Hallgreen C, Wang N, *et al*. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. *Pharmacoepidemiol Drug Saf* 2014; **23**(7): 667–78.
17 McAuslane N, Leong J, Liberti L, Walker S. The benefit–risk assessment of medicines: experience of a consortium of medium-sized regulatory authorities. *Ther Innov Regul Sci* 2017; **51**(5): 635–44.
18 Pignatti F, Ashby D, Brass E, *et al*. Structured frameworks to increase the transparency of the assessment of benefits and risks of medicines: current status and possible future directions. *Clin Pharmacol Ther* 2015; **98**(5): 522–33.
19 MDIC. A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web.pdf (accessed April 25, 2019).
20 EMA. Benefit-Risk Methodology Project Work Package 1 Report. http://eprints.lse.ac.uk/64593/1/Benefit%20risk%20methodology%20project.pdf (accessed April 25, 2019).
21 EMA. Benefit-Risk Methodology Project Work Package 2 Report. http://eprints.lse.ac.uk/64625/1/Work%20package%202%20report.pdf (accessed April 25, 2019).
22 Guo JJ, Pandey S, Doyle J, *et al*. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy – Report of the ISPOR Risk-Benefit Management Working Group. *Value Health* 2010; **13**(5): 657–66.
23 Mt-Isa S, Ouwens M, Robert V, *et al*. Structured benefit-risk assessment: a review of key publications and initiatives on frameworks and methodologies. *Pharm Stat* 2016; **15**(4): 324–32.
24 EMA. The patient's voice in the evaluation of medicines. www.ema.europa.eu/en/documents/report/report-workshop-patients-voice-evaluation-medicines_en.pdf (accessed April 25, 2019).
25 FDA. The Voice of the Patient: A Series of Reports from FDA's Patient-Focused Drug Development Initiative. www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm368342.htm (accessed April 25, 2019).
26 Hoos A, Anderson J, Boutin M, *et al*. Partnering with patients in the development and lifecycle of medicines: a call for action. *Ther Innov Regul Sci* 2015; **49**(6): 929–39.
27 Anderson M, Kimberly McCleary K. On the path to a science of patient input. *Sci Transl Med* 2016; **8**(336): 336ps11.

28 FDA. Benefit-Risk Assessment in Drug Regulatory Decision-Making. Draft PDUFA VI Implementation Plan (FY 2018–2022). www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM602885.pdf (accessed April 25, 2019).

29 Soekhai V, Donkers B, Levitan B, *et al*. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today* 2019; **24**(7): 1324–31.

30 Ryan M, Scott D, Reeves C, *et al*. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess* 2001; **5**(5): 1–186.

31 Johnson FR, Zhou M Quantifying patient preferences for regulatory benefit-risk assessments. In: Jiang Q, He W, eds. *Benefit-Risk Assessment Methods in Medical Product Development*. Boca Raton: CRC Press, 2016, pp. 85–104.

32 Muhlbacher A, Johnson FR. Choice experiments to quantify preferences for health and healthcare: state of the practice. *Appl Health Econ Health Policy* 2016; **14**(3): 253–66.

33 Bridges JF, Kinter E, Kidane L, *et al*. Things are looking up since we started listening to patients: trends in the application of conjoint analysis in health 1982–2007. *Patient* 2008; **1**(4): 273–82.

34 Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014; **4**: MR000034.

35 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342**(25): 1887–92.

36 Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ* 2016; **352**: i493.

37 Zahl PH, Maehlen J. Bias in observational studies of the association between menopausal hormone therapy and breast cancer. *PLoS One* 2015; **10**(5): e0124076.

38 Tan O, Harman SM, Naftolin F. What can we learn from design faults in the Women's Health Initiative randomized clinical trial? *Bull NYU Hosp Jt Dis* 2009; **67**(2): 226–9.

39 Bhupathiraju SN, Grodstein F, Rosner B, *et al*. Hormone therapy use and risk of chronic disease in the Nurses' Health Study: a comparative analysis with the Women's Health Initiative. *Am J Epidemiol* 2017; **186**(6): 696–708.

40 Kirkwood CW. *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Belmont: Duxbury Press, 1996.

41 EMA. Benefit-Risk Methodology Project Work Package 4 Report. www.ema.europa.eu/en/documents/report/benefit-risk-methodology-project-work-package-4-report-benefit-risk-tools-processes_en.pdf (accessed April 25, 2019).

42 Department for Communities and Local Government. *Multi-Criteria Analysis: A Manual*. http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf (accessed April 25, 2019).

43 Brass EP, Lofstedt R, Renn O. Improving the decision-making process for nonprescription drugs: a framework for benefit-risk assessment. *Clin Pharmacol Ther* 2011; **90**(6): 791–803.

44 Schulman S, Angeras U, Bergqvist D, *et al*. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in surgical patients. *J Thromb Haemost* 2010; **8**(1): 202–4.

45 Connolly SJ, Ezekowitz M, Yusuf S, *et al*. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009; **361**(12): 1139–51.

46 Irony TZ. The "utility" in composite outcome measures: measuring what is important to patients. *JAMA* 2017; **318**(18): 1820–1.

47 Kaul S, Diamond GA. Trial and error. How to avoid commonly encountered limitations of published clinical trials. *J Am Coll Cardiol* 2010; **55**(5): 415–27.

48 Levitan B, Cross J. Pharmaceutical benefit-risk assessment in early development. In: Sashegyi A, Felli J, Noel R, eds. *Benefit-Risk Assessment in Pharmaceutical Research and Development*. Boca Raton: CRC Press, 2014, pp. 3–19.

49 Colopy MW, Damaraju C, He W, *et al*. Benefit-risk evaluation and decision making: some practical insights. *Ther Innov Regul Sci* 2015; **49**(3): 425–33.

50 Unger EF. Weighing benefits and risks – the FDA's review of prasugrel. *N Engl J Med* 2009; **361**(10): 942–5.

51 Beasley BN, Unger EF, Temple R. Anticoagulant options – why the FDA approved a higher but not a lower dose of dabigatran. *N Engl J Med* 2011; **364**(19): 1788–90.

52 Bakerly ND, Woodcock A, New J, *et al*. The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in chronic obstructive pulmonary disease. *Respir Res* 2015; **16**: 101.

53 Collier S, Harvey C, Brewster J, *et al*. Monitoring safety in a phase III real-world effectiveness trial: use of novel methodology in the Salford Lung Study. *Pharmacoepidemiol Drug Saf* 2017; **26**(3): 344–52.

54 Lagakos SW. Time-to-event analyses for long-term treatments – the APPROVe trial. *N Engl J Med* 2006; **355**(2): 113–17.

55 Elemraid MA, Muller M, Spencer D, *et al*. Accuracy of the interpretation of chest radiographs for the diagnosis of paediatric pneumonia. *PLoS One* 2014; **9**(8): e106051.

56 Hazir T, Nisar Y, Qazi S, *et al*. Chest radiography in children aged 2–59 months diagnosed with non-severe pneumonia as defined by World Health Organization: descriptive multicentre study in Pakistan. *BMJ* 2006; **333**(7569): 629.

57 Ben Shimol S, Dagan R, Givon-Lavi N, *et al*. Evaluation of the World Health Organization criteria for chest radiographs for pneumonia diagnosis in children. *Eur J Pediatr* 2012; **171**(2): 369–74.

58 Graham DJ, Reichman M, Wernecke M, *et al*. Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. *Circulation* 2015; **131**(2): 157–64.

59 Makam RCP, Hoaglin D, McManus D, *et al*. Efficacy and safety of direct oral anticoagulants approved for cardiovascular indications: systematic review and meta-analysis. *PLoS One* 2018; **13**(5): e0197583.

60 Sterne JA, Bodalia P, Bryden P, *et al*. Oral anticoagulants for primary prevention, treatment and secondary prevention of venous thromboembolic disease, and for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis and cost-effectiveness analysis. *Health Technol Assess* 2017; **21**(9): 1–386.

61 Cunningham A, Stein C, Chung C, *et al*. An automated database case definition for serious bleeding related to oral anticoagulant use. *Pharmacoepidemiol Drug Saf* 2011; **20**(6): 560–6.

62 Singer DE, Chang Y, Fang M, *et al*. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med* 2009; **151**(5): 297–305.

63 Friberg L, Rosenqvist M, Lip GY. Net clinical benefit of warfarin in patients with atrial fibrillation: a report from the Swedish atrial fibrillation cohort study. *Circulation* 2012; **125**(19): 2298–307.

64 Gangireddy SR, Halperin J, Fuster V, *et al*. Percutaneous left atrial appendage closure for stroke prevention in patients with atrial fibrillation: an assessment of net clinical benefit. *Eur Heart J* 2012; **33**(21): 2700–8.

65 Gibson CM, Levitan B, Murphy S, *et al*. A net clinical outcome analysis comparing fatal or irreversible ischemic and bleeding events in

ATLAS ACS 2 - TIMI 51. *Circulation* 2012; **126**(suppl 21): A13152.

66 Banerjee A, Lane D, Torp-Pedersen C, *et al*. Net clinical benefit of new oral anticoagulants (dabigatran, rivaroxaban, apixaban) versus no treatment in a 'real world' atrial fibrillation population: a modelling analysis based on a nationwide cohort study. *Thromb Haemost* 2012; **107**(3): 584–9.

67 Chuang-Stein C, Mohberg NR, Sinkula MS. Three measures for simultaneously evaluating benefits and risks using categorical data from clinical trials. *Stat Med* 1991; **10**(9): 1349–59.

68 Rituxan Hycela(R). www.accessdata.fda.gov/drugsatfda_docs/label/2017/761064s000lbl.pdf (accessed April 25, 2019).

69 Trotti A, Colevas A, Setser A, *et al*. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* 2003; **13**(3): 176–81.

70 Guyatt GH, Akl E, Crowther M, *et al*. Introduction to the ninth edition: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012; **141**(2 Suppl): 48S–52S.

71 Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. *JAMA* 2008; **300**(4): 436–8.

72 Yuan Z, Levitan B, Burton P, *et al*. Relative importance of benefits and risks associated with antithrombotic therapies for acute coronary syndrome: patient and physician perspectives. *Curr Med Res Opin* 2014; **30**(9): 1733–41.

73 Muhlbacher AC, Juhnke C. Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? *Appl Health Econ Health Policy* 2013; **11**(3): 163–80.

74 Hauber AB, Tunceli K, Yang J, *et al*. A survey of patient preferences for oral antihyperglycemic therapy in patients with type 2 diabetes mellitus. *Diabetes Ther* 2015; **6**(1): 75–84.

75 Wilke T, Bauer S, Mueller S, *et al*. Patient preferences for oral anticoagulation therapy in atrial fibrillation: a systematic literature review. *Patient* 2017; **10**(1): 17–37.

76 Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy* 2013; **11**(4): 319–29.

77 Postmus D, Richard S, Bere N, *et al*. Individual trade-offs between possible benefits and risks of cancer treatments: results from a stated preference study with patients with multiple myeloma. *Oncologist* 2018; **23**(1): 44–51.

78 Wilson L, Loucks A, Bui C, *et al*. Patient centered decision making: use of conjoint analysis to determine risk-benefit trade-offs for preference sensitive treatment choices. *J Neurol Sci* 2014; **344**(1-2): 80–7.

79 Okumura K, Inoue H, Yasaka M, *et al*. Comparing patient and physician risk tolerance for bleeding events associated with anticoagulants in atrial fibrillation – evidence from the United States and Japan. *Value Health Reg Issues* 2015; **6**(May): 65–72.

80 Maeda K, Sakai T, Hira K, *et al*. Physicians' attitudes toward anticoagulant therapy in patients with chronic atrial fibrillation. *Intern Med* 2004; **43**(7): 553–60.

81 Hughes D, Waddingham E, Mt-Isa S, *et al*. Recommendations for benefit-risk assessment methodologies and visual representations. *Pharmacoepidemiol Drug Saf* 2016; **25**(3): 251–62.

82 Evans SR, Pennello G, Pantoja-Galicia N, *et al*. Benefit-risk Evaluation for Diagnostics: A Framework (BED-FRAME). *Clin Infect Dis* 2016; **63**(6): 812–17.

83 O'Neill RT. A perspective on characterizing benefits and risks derived from clinical trials: can we do more? *Drug Inform J* 2007; **42**: 235–45.

84 Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: examples using quantitative methods.

Pharmacoepidemiol Drug Saf 2003; **12**(8): 693–7.

85 Schnipper LE, Davidson N, Wollins D, *et al*. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. *J Clin Oncol* 2015; **33**(23): 2563–77.

86 Fischhoff B, Brewer NT, Downs JS. *Communicating Risks and Benefits: An Evidence-Based User's Guide*. Washington DC: FDA/ DHHS, 2011.

87 Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999; **319**(7223): 1492–5.

88 Hildebrandt M, Vervolgyi E, Bender R. Calculation of NNTs in RCTs with time-to-event outcomes: a literature review. *BMC Med Res Methodol* 2009; **9**: 21.

89 Holden WL.,Benefit-risk analysis: a brief review and proposed quantitative approaches. *Drug Saf* 2003; **26**(12): 853–62.

90 Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998; **317**(7168): 1309–12.

91 Bender R. Calculating confidence intervals for the number needed to treat. *Control Clin Trials* 2001; **22**(2): 102–10.

92 Fang MC, Go A, Chang Y, *et al*. Death and disability from warfarin-associated intracranial and extracranial hemorrhages. *Am J Med* 2007; **120**(8): 700–5.

93 Rivaroxaban Cardiovascular and Renal Drugs Advisory Committee Briefing Document March 19, 2009. https://wayback.archive-it.org/7993/20170405212708/www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM138385.pdf (accessed April 25, 2019).

94 Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Med Decis Making* 2009; **29**(1): 104–15.

95 Hsu JC, Hsieh C, Yang Y, *et al*. Net clinical benefit of oral anticoagulants: a multiple criteria decision analysis. *PLoS One* 2015; **10**(4): e0124806.

96 Phillips LD, Fasolo B, Zafiropoulous N, *et al*. Modelling the risk-benefit impact of H1N1 influenza vaccines. *Eur J Public Health* 2013; **23**(4): 674–8.

97 Nixon R, Nguyen T, Stoeckert I, *et al*. IMI WP5 Report 1:b:iv Benefit-Risk Wave 1 case study report: NATALIZUMAB. www.imi-protect.eu/documents/NixonetalBenefitRiskWave1casestudyreportNatalizumabMay2013.pdf (accessed April 25, 2019).

98 Hallgreen CE, van den Ham H, Mt-Isa S, *et al*. Benefit-risk assessment in a post-market setting: a case study integrating real-life experience into benefit-risk methodology. *Pharmacoepidemiol Drug Saf* 2014; **23**(9): 974–83.

99 Waddingham E, Mt-Isa S, Nixon R, *et al*. A Bayesian approach to probabilistic sensitivity analysis in structured benefit-risk assessment. *Biom J* 2016; **58**(1): 28–42.

100 Juhaeri J, Amzal B, Chan E, *et al*. IMI Work Package 5: Report 2:b:i Benefit - Risk Wave 2 Case Study Report: Rimonabant. www.imi-protect.eu/documents/JuhaerietalBenefitRiskWave2CaseStudyReportRimonabantJan2012.pdf (accessed April 25, 2019).

101 Hauber AB, Johnson F, Fillit H, *et al*. Older Americans' risk-benefit preferences for modifying the course of Alzheimer disease. *Alzheimer Dis Assoc Disord* 2009; **23**(1): 23–32.

102 Katz EG, Hauber B, Gopal S, *et al*. Physician and patient benefit-risk preferences from two randomized long-acting injectable antipsychotic trials. *Patient Prefer Adherence* 2016; **10**: 2127–39.

103 Hallgreen CE, Mt-Isa S, Lieftucht A, *et al*. Literature review of visual representation of the results of benefit-risk assessments of medicinal products. *Pharmacoepidemiol Drug Saf* 2016; **25**(3): 238–50.

104 Wen, S, *et al*. Visualizaton of benefit-risk assessment in medical products with real examples. In: Jiang Q, He W, eds. *Benefit-Risk Assessment Methods in Medical Product Development*. Boca Raton: CRC Press, 2016, pp. 197–231.

105 21st Century Cures Act. 2016. www.congress.gov/bill/114th-congress/house-bill/34/text (accessed April 25, 2019).

106 MDUFA Performance Goals And Procedures, Fiscal Years 2018 Through 2022. www.fda.gov/downloads/ForIndustry/UserFees/MedicalDeviceUserFee/UCM535548.pdf (accessed April 25, 2019).

107 Jiang Q, He W. *Benefit-Risk Assessment Methods in Medical Product Development: Bridging Qualitative and Quantitative Assessments*. Boca Raton: CRC Press, 2016.

108 Peay HL, Hollin I, Fischer R, *et al*. A community-engaged approach to quantifying caregiver preferences for the benefits and risks of emerging therapies for duchenne muscular dystrophy. *Clin Ther* 2014; **36**(5): 624–37.

109 Uno H, Wittes J, Fu H, *et al*. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Ann Intern Med* 2015; **163**(2): 127–34.

110 Evans S, Ting N. *Fundamental Concepts for New Clinical Trialists*. Boca Raton: CRC Press, 2016.

111 DeMets DL, Califf RM. A historical perspective on clinical trials innovation and leadership: where have the academics gone? *JAMA* 2011; **305**(7): 713–14.

112 Norton JD, Longitudinal model and graphic for benefit-risk analysis, with case study. *Drug Inform J* 2011; **45**: 741–7.

113 Zhao Y, Zalkikar J, Tiwari R, *et al*. A Bayesian approach for benefit-risk assessment. *Stat Biopharm Res* 2014; **6**(4): 326–77.

114 Costa MJ, He W, Jemiai Y. The case for a Bayesian approach to benefit-risk assessment: overview and future directions. *Ther Innov Regul Sci* 2017; **51**(5): 568–74.

115 Evans SR, Rubin D, Follmann D, *et al*. Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR). *Clin Infect Dis* 2015; **61**(5): 800–6.

116 Evans SR, Follmann D. Using outcomes to analyze patients rather than patients to analyze outcomes: a step toward pragmatism in benefit:risk evaluation. *Stat Biopharm Res* 2016; **8**(4): 386–93.

117 Van Duin D, Lok J, Earley M, *et al*. Colistin versus ceftazidime-avibactam in the treatment of infections due to carbapenem-resistant enterobacteriaceae. *Clin Infect Dis* 2018; **66**(2): 163–71.

118 Bonetti M, Gelber RD. Patterns of treatment effects in subsets of patients in clinical trials. *Biostatistics* 2004; **5**(3): 465–81.

119 Pennello G, Pantoja-Galicia N, Evans S. Comparing diagnostic tests on benefit-risk. *J Biopharm Stat* 2016; **26**(6): 1083–97.

120 Evans S, Liu Y, Follmann D. Pragmatic Benefit:Risk Evaluation: Healthy Disruption for Clinical Trials. www.bidmc.org/-/media/files/beth-israel-org/smith-center-for-outcomes-research-in-cardiology/10-24.ashx?la=en&hash=7449D33BD333D22F9B958A98C5D9CA08F2D0798A (accessed April 25, 2019).

36

The Use of Metaanalysis in Pharmacoepidemiology

Jesse A. Berlin[1], Brenda J. Crowe[2], H. Amy Xia[3], and Stephen J.W. Evans[4]

[1] *Johnson & Johnson, Titusville, NJ, USA*
[2] *Eli Lilly & Co., Indianapolis, IN, USA*
[3] *Amgen, Inc., Thousand Oaks, CA, USA*
[4] *London School of Hygiene & Tropical Medicine, London, UK*

Definitions and Rationale

Metaanalysis has been defined as "the statistical analysis of a collection of analytic results for the purpose of integrating the findings" [1]. Other definitions have included qualitative, as well as quantitative, analyses [2]. Metaanalysis is used to identify sources of variation among study findings and, when appropriate, to provide an overall measure of effect as a summary of those findings [3]. A systematic review has been defined by the Cochrane Group (https://consumers.cochrane.org/what-systematic-review) as a review that "summarizes the results of available carefully designed healthcare studies (controlled trials) and provides a high level of evidence on the effectiveness of healthcare interventions." This definition illustrates two features related to the Cochrane Collaboration; firstly, their emphasis on randomized trials and secondly, their emphasis on effectiveness. All good science seeks to take into account all the evidence related to a question of interest, and pharmacoepidemiology seeks to be a good science. There are two main processes involved in taking into account all evidence: identifying all the evidence, and providing a useful summary. This chapter concentrates on summarizing the evidence using statistical methods but the need to search the literature is usually also important. The term "metaanalysis" may be used synonymously with "systematic review" but here we employ the definition used in the report of the Council for International Organizations of Medical Sciences (CIOMS) Working Group X [4]:

> The statistical combination of quantitative evidence from two or more studies to address common research questions, where the analytical methods appropriately take into account that the data are derived from multiple individual studies.

Metaanalysis in medicine has been developed mainly for combination of data from randomized trials, but this chapter also includes evidence from observational studies. Metaanalyses themselves could be regarded as observational studies and should perhaps be simply seen as a technique for combining evidence rather than as possessing particular merit as studies in themselves. The *Journal of the American Medical Association*, in a statement by the Editors, goes

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

even further: "JAMA considers meta-analysis to represent an observational design, such that outcomes, inferences, and interpretations should be described as associations rather than reported using causal terms such as 'size of effect' …" [5].

While epidemiologists have been cautious about adopting metaanalysis, because of the inherent biases in the component studies and the great diversity in study designs and populations [6–8], the need to make the most efficient and intelligent use of existing data prior to (or instead of) embarking on a large, primary data collection effort has dictated a progressively more accepting approach [8–13]. Metaanalysis of randomized clinical trials has found such wide acceptance that the Cochrane Collaboration has been built around the performance and updating of systematic reviews and metaanalyses of trials [14]. Cochrane reviews are maintained in a publicly available electronic library. More information is available on the Cochrane website (www.cochrane.org). A similar structure has developed in the social sciences, in the form of the Campbell Collaboration [15].

As the Cochrane Collaboration has increasingly looked at harms as well as benefits of health interventions, it has been realized that observational studies may make an important contribution, and that overall they are not as biased as some think [16]. The Cochrane Adverse Effects Method Group (AEMG) is focused on systematic reviews related to adverse effects (https://methods.cochrane.org/adverseeffects/about-caemg) [17] and this group notes that observational data are needed in some circumstances. The current version of the Cochrane Handbook [18] has chapters on "Including nonrandomized studies" and "Adverse effects." In the introduction on nonrandomized studies, they are now quite explicit:

> For some Cochrane reviews, the question of interest cannot be answered by randomized trials, and review authors may be justified in including non-randomized studies.

Several activities may be included under the above definition of metaanalysis. Perhaps the most popular conception of metaanalysis, for most clinically oriented researchers, is the summary of a group of randomized clinical trials dealing with a particular therapy for a particular disease. An example of this approach would be a metaanalysis that examined the effects of antiplatelet therapy in high-risk patients [19]. Typically, this type of metaanalysis would present an overall measure of the efficacy of treatment, such as a summary odds ratio. Summary measures may be presented for different subsets of trials involving specific types of patients, such as studies restricted to those with a previous myocardial infarction. More sophisticated metaanalyses also examine the variability of results among trials and, when results have been conflicting, attempt to uncover the sources of the disagreements [20].

In many instances, major collaborations have been set up to reanalyze the data at an individual level rather than just relying on published summary data. This "individual-patient-data" (IPD) approach is useful with both randomized and nonrandomized studies.

Metaanalyses of nonexperimental epidemiologic studies have also been performed [21–24], and articles have been written describing the methodologic considerations specific to those metaanalyses [25–31]. In general, both the metaanalyses of nonexperimental studies and the associated methodologic articles tend to focus more on the exploration of reasons for disagreement among the results of prior studies, including the possibility of bias. Given the greater diversity of designs and heterogeneity of patients included in nonexperimental studies, it is logical to find more disagreement among nonexperimental studies than among randomized trials.

Metaanalysis of randomized trials frequently has increased precision as the main motivation, especially in the context of safety of medicines where outcomes are rare. The sample sizes in

the original trials are often far too small to provide convincing evidence of the presence or absence of harms. This is not the same motivation in observational studies in general. They will have been undertaken with, usually, dramatically larger sample sizes so the issue of precision of estimation is often not at the forefront. The consequence is that the problems of uncontrolled bias can have a very much larger impact than the size of effect being studied. Sampling variation, in contrast to randomized trials, is not the main problem. Consequently, the danger is that such studies will have minimal sampling uncertainty but unknown bias.

This chapter summarizes many of the major conceptual and methodologic issues surrounding metaanalysis, especially for observational studies, and offers the views of the authors about possible avenues for future research in this field.

Clinical Problems to be Addressed by Pharmacoepidemiologic Research

There are a number of reasons why a pharmacoepidemiologist might be interested in conducting a metaanalysis. If randomized data are available, although each study allows for comparisons relatively free from the confounding and bias of nonexperimental studies, individual studies will often have insufficient power to detect as significant uncommon adverse outcomes of therapies that nevertheless can make a notable alteration in the risk/benefit balance. In addition, the exploration of reasons for inconsistencies of results across previous studies, the exploration of subgroups of patients in whom therapy may be more or less effective, and understanding whether risk factors affect the magnitude of treatment differences are all valid topics for study. The combination of studies involved in the approval process for new therapies or in the investigation of new indications for existing therapies, particularly when the outcomes being studied are uncommon, is another area where metaanalysis can be useful.

With the investigation of adverse events using nonexperimental studies, a major challenge involves obtaining information on these events that is unconfounded by indication (see Chapter 37). These adverse events often occur only rarely and will not have standardized assessment or validation, which makes their evaluation still more difficult. The results of nonexperimental studies of whether such events are associated with a particular drug may be conflicting, leaving a confusing picture for practicing clinicians and policy makers to interpret. Metaanalysis, by combining results from many *randomized* studies, may be better than relying on nonexperimental studies with their potential confounding and bias. When reports of several investigations of a specific suspected adverse drug reaction disagree, whether randomized or nonexperimental in design, metaanalysis can also be used to help explain these disagreements. These disagreements among studies may arise from differences in the choice of endpoints, the exact definition of exposure, the eligibility criteria for study subjects, the methods of obtaining information, differences in protocols, different methods to analyze the results or a host of other reasons possibly related to the susceptibility of the constituent studies to bias.

While it is not possible to produce a definitive answer to every research question, the exploration of the reasons for heterogeneity among study results may at least provide valuable guidance concerning the design of future studies. For example, separating those studies that had the possibility of recall bias in relation to drug exposure from those that did not, when examining associations of antihypertensive drugs used in mothers and congenital anomalies in their offspring [32], showed that cardiovascular anomalies were associated with drugs in studies

possibly subject to recall bias but not in those studies with ascertainment of exposure prior to birth.

Historically in drug development, it has been common practice to look at safety data from individual trials in isolation as they are completed. Subsequently, just prior to submission of a new product application, data from multiple studies are summarized in an "Integrated Summary of Safety" or "Summary of Clinical Safety" report. A possible consequence of these practices is that the opportunity to respond earlier to the evolving safety and tolerability profile (by collecting additional data or adjusting the sample size of pivotal studies, for example) may be missed. The result might be a gap (that might have been avoided) in the knowledge of the safety profile at the time of submission, and this may generate further questions by regulatory agencies or prompt the need for additional postmarketing commitments. In addition, the practice of simply using "crude pooling" can be misleading. This is the simple adding up across all studies of the numbers of patients with an event in the treatment group and the total number of patients receiving that treatment, and doing the same for the control group to produce a single 2 × 2 table which is then analyzed ignoring the fact that the data came from different trials (studies). This is often biased, unless the ratio of patients receiving the new drug to that receiving the comparator is approximately the same across all studies included in the integrated analysis, or the incidence of the event is nearly the same across all studies in the comparator group [33].

In cases where the individual studies require methods that take time into account (such as studies where the follow-up/exposure time among the treatment groups is substantially different), methods that take time into account are necessary. There may also be differences in length of follow-up across studies (independent of any differences in follow-up within studies). For a more in-depth discussion of issues related to follow-up time, we recommend Section 3.4 in the CIOMS X report [4].

Industry and regulatory agencies are placing increasing emphasis on identifying safety signals for new compounds early in the drug development process. As a response, the Safety Planning, Evaluation and Reporting Team (SPERT) was formed in 2006 by the Pharmaceutical Research and Manufacturers of America (PhRMA). The goal of SPERT was to propose a standard across the pharmaceutical industry for safety planning, data collection, evaluation, and reporting, beginning with planning first-in-human studies and continuing through the planning of postapproval activities [34].

Among the key recommendations from SPERT was that sponsors plan a series of repeated, cumulative metaanalyses of the safety data obtained from the studies conducted within the development program. Leading up to these metaanalyses, sponsors need to develop clear definitions of adverse events of special interest and standardize various aspects of data collection and study design, to facilitate combining studies and the interpretation of the combined analyses. These ideas were extended in the CIOMS X Working Group report. It is important to ensure that true adverse reactions to medicines and vaccines are known about so that clinicians and patients can be aware and take appropriate action, but it is equally important to show that the magnitude of any risks in absolute terms is such that the benefit/risk balance for a medicine in a particular situation remains positive. It is never possible to show with absolute certainty the absence of a possible harm since the confidence interval will always include such a possibility unless (assuming a well-designed study) the confidence interval excludes zero risk (in the direction of a benefit), in which case the adverse effect is prevented and is no longer a harm for that medicine.

By following a proactive approach during development, including periodic updating of cumulative metaanalyses, potential harms may

be identified earlier in the development process. This may increase the chances that the Phase III program will be able to provide a satisfactory understanding of the safety profile. Risk management plans will take into account the knowledge about the drug at the time of licensing (see Chapter 24) [35]. Furthermore, the needs for postmarketing commitments can be better defined.

The exploration of subgroups of patients in whom therapy may be more or less effective is a controversial question in individual randomized trials. Most trials are not designed with sample sizes adequate to address efficacy in subgroups. The finding of statistically significant differences between the effects of therapy in different subgroups, particularly when those groups were not defined *a priori*, raises the question of whether those are spurious findings. Conversely, the lack of statistical significance for clinically important differences between prospectively defined subgroups can often be attributed to a lack of statistical power. Such clinically meaningful but statistically nonsignificant findings are difficult to interpret. Metaanalysis can be used to explore these questions with improved statistical power. A prespecified protocol is important, especially in this context, since the potential for adjusting methods to find or not find effects can be dependent on the interests of those doing the metaanalysis.

The use of metaanalysis in the approval process for new drugs or devices represents another potential application, although experience in this area is as yet rather limited. However, many of the methodologic issues arising in the context of new drug approval also arise in the investigation of new indications for pharmaceutical products that have previously been approved for other purposes. For some therapies, such as streptokinase in the treatment of myocardial infarction, metaanalysis could have been used to summarize evidence prior to embarking on a very large-scale, multicenter, randomized trial [36].

Evidence-based medicine requires the use of the best evidence available in making decisions about the care of patients. Traditional metaanalyses, which have been one of the cornerstones of evidence-based medicine, often focus on placebo-controlled trials because head-to-head comparisons of medications are generally unavailable. But what healthcare providers, patients, and policy makers need to make better informed decisions is an analysis that provides comprehensive look at all available evidence – how a specific pharmacologic treatment compares with other available pharmacologic treatments in terms of safety and efficacy for the specific condition (see Chapter 32).

Extended metaanalytic techniques such as indirect comparisons [37] and multiple treatment metaanalyses can combine all available evidence in a single analysis [38]. These techniques provide estimates of the effect of each intervention relative to every other, whether or not they have been directly compared in trials, allowing ranking of treatments in terms of efficacy and safety, and can potentially strengthen the inference regarding a treatment because the results are based on more data. The main drawback of these analyses is that the validity of the findings depends on whether homogeneity and consistency assumptions, which we describe later, are met [39]. There has been a growth in "network metaanalysis" (NMA, also known as mixed treatment comparisons), which is another phrase to describe combining studies of multiple treatments allowing for direct and indirect comparisons and assessing their consistency. A helpful website is www.bristol.ac.uk/population-health-sciences/centres/cresyda/mpes/mtc/. There is more discussion of this in the section "Indirect comparison and simultaneous comparison of treatments" later. This technique can address the clinical problem when it is desired to make comparisons between active treatments but there are no or very few trials that make the direct comparisons.

Methodologic Problems to be Solved by Pharmacoepidemiologic Research

As the skeptical reader might imagine, many methodologic issues can arise in the context of performing a metaanalysis. Many, but not all, of these problems relate to the process of combining studies that are often diverse with respect to specific aspects of design or protocol, some of which may be of questionable quality.

Susceptibility of the Original Studies to Bias

Early work in metaanalysis used the term "study quality." More recent efforts (e.g., PRISMA [40]) have adopted language that refers to susceptibility to bias or likelihood of bias. We adopt that new terminology in the remainder of this chapter. The Cochrane Collaboration developed a "risk of bias tool" which is now used extensively (at least with many thousands of citations in the literature) in assessing the randomized trials to be included in a metaanalysis [41]. This paper describes some evaluation, and there is a chapter in the Cochrane Handbook that gives many more details [18].

The tool in use has been developed further and there are training materials available freely on the internet for their use (https://training.cochrane.org/resource/rob-20-webinar). A further development is the ROBIS tool which, rather than assessing the individual trials, assesses the risk of bias in a systematic review or metaanalysis itself [42]. The group that developed ROBIS has also developed a similar approach to risk of bias in observational studies, ROBINS-I (Risk Of Bias In Nonrandomized Studies – of Interventions) [43] but this has been less heavily cited and newer versions are under development.

As might be expected, the Cochrane groups are not the only ones which have developed such tools, but they have published a systematic review of such tools [44]. They conclude:

> There are several limitations of existing tools for assessing risk of reporting biases, in terms of their scope, guidance for reaching risk of bias judgments and measurement properties. Development and evaluation of a new, comprehensive tool could help overcome present limitations.

It is unrealistic to expect complete agreement on a single method of assessing bias but it is clearly an active area, and there is an obvious need for those doing metaanalysis in pharmacoepidemiology to be aware and to ensure that at least some attempts are made to assess bias in more than a trivial way.

Metaanalysis seems particularly prone to the "garbage in = garbage out" phenomenon. Combining a group of poorly done studies can produce a precise summary result built on a very weak foundation. This apparent precision may lend undue credibility to a result that truly should not be used as a basis for formulating clinical or policy strategies [7]. However, if the judgment about susceptibility to bias in an individual study is subtly influenced by the direction or magnitude of the findings of the study, excluding studies based on such a subjective judgment about their quality could open the metaanalytic process to a different, and potentially serious, form of bias.

Combinability of Studies

Different studies will look at different outcomes, with different treatments in different patients. Some studies will look at multiple outcomes, treatments or groups of patients. Clearly, no one would suggest combining studies that are so diverse in either outcomes or treatments that a summary would be nonsensical. For example,

one would not combine studies of hormone replacement therapy in relation to risk of breast cancer with studies of hormone replacement therapy in relation to risk of coronary heart disease. The outcomes are not expected to show the same effect. Beyond obvious examples like this, however, the choices may not be so clear. Should studies with different patient populations be combined? How different can those populations be before it becomes unacceptable to combine the studies? Should different hormone preparations be considered as being equivalent (those with estrogen only, those with estradiol, etc.)? It will require judgment; for some purposes it may be relevant to look at all hormone preparations as a single group, but for other purposes it will be important to distinguish products, or even different doses and durations of treatment for a single product. In pharmacoepidemiology, risks of adverse effects will vary notably with duration of treatment and length of follow-up. Careful consideration needs to be applied before combining studies with diverse outcome definitions and methodology. At the very least, some exploration of the sources of variation of effect should be done.

As an example where combining studies may be appropriate, a recent metaanalysis, Sabatine *et al*. [45], combines across all statins but looks at a very restricted population of patients with low LDL levels at baseline. In this context, it may be reasonable to combine different statins used under different conditions in order to see if effects are present in a group with already low LDL levels.

Should nonrandomized studies be combined with randomized studies? Should nonrandomized studies ever be used in a metaanalysis? Should studies with active drugs as comparators be combined with studies with placebos as comparators? These are questions that cannot be answered without generating some controversy. The CIOMS Working Group X concluded in regard to the first question: "A clear consequence of this is that the treatment of randomized and non-randomized evidence as equivalent (exchangeable) in a single analysis is a mistake; combination of data must distinguish between them" [4]. However, they go on to note: "Some efforts have been made, particularly in a Bayesian paradigm, to combine evidence from both sources in a single analysis, but the methods for doing this are not yet widely agreed." There would be little argument that they are not equivalent and reporting the results separately should be done, even if combined at some point.

Publication Bias

Unpublished material cannot be retrieved by literature searches and is likely to be difficult to find referenced in published articles. Publication bias occurs when study results are not published, or their publication is delayed, because of the results [46–56]. The usual pattern is that statistically significant results are more likely to be published than nonsignificant results, although this bias may not be as severe for randomized studies as it is for nonrandomized studies [48,57,58]. With the introduction of registration of randomized trials, publication bias in randomized studies should become less of a problem. As noted by the Office for Human Research Protections in the US Department of Health & Human Services, the International Committee of Medical Journal Editors announced a policy in 2004 that as a condition of publication, clinical trials would be required to be listed in a public registry [59]. Subsequently, regulatory authorities around the world began to require the posting of clinical trial information and, in some cases, the submission of summary results to a publicly accessible registry. Likewise, some research funding agencies are now encouraging or requiring the registration and results reporting of the clinical trials they fund. They go on to list three separate international registries as well as the best-known Clinicaltrials.gov from the US, a Canadian one,

five in Europe, and 16 from elsewhere in the world (www.hhs.gov/ohrp/international/clinical-trial-registries/index.html). Some of these registers also contain observational study protocols and results; notable is the register of Post-authorization Studies (PAS) maintained at the European Medicines Agency by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP®) (www.encepp.eu/encepp/studiesDatabase.jsp).

While one could simply decide not to include unpublished studies in a metaanalysis, since those data have often not been peer reviewed [60], unpublished data can represent a large proportion of all available data [61]. Published results are a biased sample – as noted in the Cochrane Handbook (Section 10.1 Introduction) [18]:

> Statistically significant, "positive" results that indicate that an intervention works are more likely to be published, more likely to be published rapidly, more likely to be published in English, more likely to be published more than once, more likely to be published in high impact journals and, related to the last point, more likely to be cited by others. The contribution made to the totality of the evidence in systematic reviews by studies with non-significant results is as important as that from studies with statistically significant results.

If the results of unpublished studies are systematically different from those of published studies, particularly with respect to the magnitude and/or direction of the findings, their omission from a metaanalysis would yield a biased summary estimate (assuming that the quality of the unpublished studies is at least equal to the quality of the published studies). For example, Decullier *et al.* have shown that in France, the filtration process from protocol to publication leads to more positive results appearing in the literature [62]. Hopewell *et al.* reviewed several publications and demonstrated that positive results were more likely to be published [63].

Publication bias is a potentially serious limitation to any metaanalysis. For example, Sutton and colleagues found that in four of 48 metaanalyses they examined, there was evidence that the statistical inferences would have changed after the overall effect estimate was adjusted for publication bias [64]. The retrospective identification of completed unpublished trials is clearly possible [61] in some instances, but generally is not practical. One study used a survey of investigators to attempt to identify unpublished studies [65]. The authors surveyed 42 000 obstetricians and pediatricians, asking whether they had participated in any unpublished trials completed more than two years previously, that is, during the period prior to the end of 1984. They identified only 18 such studies, despite an overall response rate of 94% to their survey.

Other forms of bias, related to publication bias, have also been identified [49]. These include reference bias, – preferential citation of significant findings [66]; language bias – exclusion of studies in languages other than English [67,68]; and bias related to source of funding [69–71]. These related biases have been termed "dissemination bias" by Sutton and colleagues, who found that the threat of such biases is more severe in nonrandomized studies of an intervention [64].

There is room to decrease publication bias even further: calls to simplify access to the US Food and Drug Administration (FDA) and other regulatory agency reviews, and to create links from such reviews to literature search engines such as MEDLINE, have been made [72]. Ladanie and colleagues have made practical suggestions regarding how to find FDA reviews [73].

Unfortunately, there is no guarantee that either the published trial or a published metaanalysis will follow the protocol. It seems that bias towards finding "interesting" results is pervasive. Hutton and Williamson showed that bias within a study in choosing outcomes to be

reported, when many have been studied, could affect the reported results and hence affect a metaanalysis [74]. Marzouki and colleagues, using the registered trial protocols at *The Lancet,* showed that in 11 out of 37 trials, there were major discrepancies in the primary outcome between those protocols and the subsequent publication [75]. Dwan *et al.* have reviewed the problem in a general way and find real causes for concern [76]. It is one thing to have the trial data biased, but then the systematic reviewers may add to the problem. Kirkham *et al.* studied protocols and subsequent publications from the Cochrane Library, which is thought by some to apply the highest standards of reviews, and reported "Over a fifth (64/288, 22%) of protocol/review pairings were found to contain a discrepancy in at least one outcome measure, of which 48 (75%) were attributable to changes in the primary outcome measure" [77]. Saini *et al.*, from the same group, using the same data source, also found that harms were particularly unreliably reported in Cochrane reviews [78].

Rising *et al.* [79] suggested that the published literature gave a more favorable impression than the FDA reviews had found, while in the narrower area of biologics, Amarilyo and Furst found discrepancies between FDA reviews and publications but there was no consistent difference in the direction [80]. It is clear that the published literature may not be as reliable as it should be and that regulatory review may be more reliable, and at the very least should also be searched for in any metaanalysis based on published data.

Bias in the Abstraction of Data

Metaanalysis, by virtue of being conducted after the data are available, is a form of retrospective research and is thus subject to the potential biases inherent in such research [81]. In a metaanalysis of gastrointestinal side effects of NSAIDs, Chalmers and colleagues examined over 500 randomized studies [82]. They measured the agreement of different reviewers when reading the "methods" sections of papers that had been masked as to their source and the results. There were disagreements on 10–20% of items, which had to be resolved in conference with a third person. These disagreements arose from errors on the part of the reader and from lack of clarity of the presentation of material in the original articles. Whatever its source, when such variability exists, the opportunity for observer bias may exist as well [81].

In a number of instances, more than one metaanalysis has been performed in the same general area of disease and treatment. A review of 20 of these instances showed that, for almost all disease/treatment areas, there were differences between two metaanalyses, of the same topic, in the acceptance and rejection of papers to be included [60]. While there was only one case (out of the 20) of extreme disagreement regarding efficacy, there were several cases in which one or more analyses showed a statistically significant result while the other(s) did not. These disagreements were not easily explainable. For example, differences between metaanalyses of the same topic in the acceptance and rejection of papers did not always lead to differences in conclusions.

More generally, the acceptance or rejection of different sets of studies can drastically change conclusions. Despite efforts to make metaanalysis an objective, reproducible activity, there is evidently some judgment involved.

In a separate commentary, DerSimonian reanalyzed data from one metaanalysis and one clinical review of parenteral nutrition with branched chain amino acids in hepatic encephalopathy [83]. She pointed to differences in the data extracted by the two sets of authors [84,85] for the same endpoints from the same original papers. When combined statistically, the data extracted by the two sets of authors led to substantively different conclusions about the efficacy of therapy.

In some instances, attempts are made to extract numerical data from figures and software exists to do this. There are suggestions that extracting data from figures with software was faster, with higher interrater reliability than manual extraction [86]. However, neither method will work for the type of large studies encountered in pharmacoepidemiology, and clearly it is better to obtain raw data from authors or repositories if possible.

There is the possibility of reducing reliance on abstraction of data by using the original data from trials or observational studies. As noted above, PRISMA has a guidance on utilizing IPD [87] and there have been a large number of metaanalyses both of randomized and nonrandomized IPD in the past. (e.g., on trials in breast cancer [88] and on hormonal factors in breast cancer [89]).

Metaanalyses based on IPD have several advantages (see, for example, Section 3.9 of the CIOMS report on metaanalysis [4] and Berlin *et al.* [90]). There are strong moves to make at least individual patient randomized trial data available, such as The Yale University Open Data Access (YODA) Project (http://yoda.yale.edu/welcome-yoda-project). "Alltrials" is an international initiative of multiple stakeholders including journals, academics, and foundations (www.alltrials.net/). European regulators have promised to make data more available (www.ema.europa.eu/docs/en_GB/document_library/Other/2013/06/WC500144730.pdf) but there are privacy barriers in some instances, and there is controversy over whether reanalysis of published data needs to have some constraints [91] or not [92].

Most of the interest in this area is around randomized trials, but there are also issues around the availability of observational data. Platt and Lieu have noted that there are challenges to overcome in spite of enthusiasm of the research community for wider availability of patient data [93]. They identify three reasons for the challenges: (1) confidentiality and proprietary concerns, (2) the cost and work required to make raw data usable for analyses, and (3) the need to create incentives for data holders that outweigh the disadvantages. Concerns over privacy led to the collapse of a major initiative in the UK to share patient data, known as "care.data" [94].

While these challenges relate to overall availability, there are similar problems in making individual records from observational studies available routinely. The major collaborations that have used individual data have had to ensure that the data are only available to named team members under strict control for confidentiality.

Currently Available Solutions

This section will first present the general principles of metaanalysis and a framework for the methods typically employed in a metaanalysis. Since much of the general framework for conducting systematic reviews and explanation of the methods typically employed in a metaanalysis have been presented in review articles in major clinical journals [10,11,95], freely accessible guidelines, and handbooks, only the most important points will be highlighted here. In this chapter, we will provide succinct descriptions of the most recent guidelines and references.

The PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses) was developed to increase the clarity and transparency of published systematic reviews and metaanalyses. It consists of a 27-item checklist and a flow diagram. It describes the rationale for including each of the items with supporting references and provides examples of good reporting, and there is a short paper published simultaneously in six journals, e.g., Moher *et al.* [96], as well as a fuller "Explanation and Elaboration" paper published in three journals [40]. The flow diagram describes the number of

studies at each phase of the metaanalysis, starting with the number of studies identified in database searching, moving to the number of studies screened, those determined to be potentially eligible, and finally the number of studies included. A similar guideline for reporting metaanalysis of observational studies is available as well [25].

A very large range of guidelines has been brought together under the "EQUATOR" (Enhancing the QUAlity and Transparency Of health Research) umbrella (www.equator-network.org/), including ones for trials (CONSORT) as well as systematic reviews (PRISMA), together with many extensions for specific topics such as harms. The PRISMA on harms is particularly relevant in pharmacoepidemiology and is, like most of the EQUATOR guidelines, continually under review [97]. There are two further guidelines of special interest to pharmacoepidemiology on network metaanalysis [98] and individual patient data [87]. An example evaluation of the use of the PRISMA guideline has been conducted in the field of gastroenterology and hepatology; Panic *et al.* claim that "The endorsement of PRISMA resulted in increase of both quality of reporting and methodological quality" [99]. However, Page and Moher, who examined all the published evaluations of the use of PRISMA in systematic reviews (SRs) they could find, concluded "Many studies have evaluated how well SRs adhere to the PRISMA Statement, and the pooled result of these suggest that reporting of many items is suboptimal. An update of the PRISMA Statement, along with a toolkit of strategies to help journals endorse and implement the updated guideline, may improve the transparency of SRs" [100].

We have cited another very useful source of information on how to conduct systematic reviews and metaanalysis, namely the Cochrane Handbook for Systematic Reviews [101]. The Handbook is a publicly available, comprehensive, and easy-to-read document that describes in detail the process of preparing a systematic review, combining data, and maintaining Cochrane systematic reviews.

In the second part of this section, specific solutions to the methodologic issues raised in the previous section are presented. Finally, case studies of applications that should be of interest to pharmacoepidemiologists will be presented, illustrating approaches to some of the clinical and methodologic problems raised earlier.

Steps Involved in Performing a Metaanalysis (Box 36.1)

The CIOMS X report suggests a metaanalysis protocol should include, but is not limited to, content implied by a series of topic headings [4]. That report elaborates on the content in the context of planning the metaanalysis. The reader is referred to the CIOMS publication for more detail. We focus here on a subset of that content.

Define the Purpose

While this is an obvious component of any research, it is particularly important to define precisely the primary and secondary objectives of a metaanalysis. A well-formulated question should have a clearly defined patient problem, intervention, comparator, and outcome of interest. This framework is called PICO which stands for patient, intervention, comparison, and outcome [102]. Some authors have expanded PICO to PICOS or PICOTS, with "T" standing for timing and "S" representing either "study design" or "setting".

Box 36.1 General steps involved in conducting a metaanalysis

1) Define purpose
2) Perform literature search
3) Establish inclusion/exclusion criteria
4) Collect the data
5) Perform statistical analysis
6) Formulate conclusions and recommendations

An example of an important primary question could be: "What is the magnitude of the increased risk of gastrointestinal side effects with NSAIDs used for the treatment of pain, compared with placebo?" Another might be: "Are corticosteroids effective in the treatment of alcoholic hepatitis, compared with placebo?" Secondary objectives might include the identification of subgroups in which a treatment appears to be uniquely more or less effective. For NSAIDs, estimating the absolute risk difference (and, thus, the public health implications) as well as the relative risk (and, thus, the etiologic implications) might be a secondary objective. We present more on the question of absolute risks, relative risks, and odds ratios later. It is important to consider that questions defined too broadly could lead to the criticism of "mixing apples and oranges" and that questions focused too narrowly could lead to finding no, or limited, data, or the inability to generalize the study results.

Perform the Literature Search (Included in "Sources of data" in CIOMS X)

While computerized searches of the literature can facilitate the retrieval of all relevant published studies, these searches are not always reliable. Several studies have examined problems with the use of electronic searches [103–105]. Use of search terms that are too nonspecific can result in large numbers of mostly irrelevant citations that need to be reviewed to determine relevance. Use of too many restrictions can result in missing a substantial number of relevant publications.

Search strategies to identify specifically reports of all definite or possible randomized or quasirandomized trials have been developed. One of these strategies is the Cochrane search strategy. Although this strategy is highly sensitive (it identifies 92% of trials), the specificity is very low (3.7%) (i.e., it identifies a lot of nonrelevant studies) [106]. Nonetheless, one term "random*[tw]" is able to retrieve all randomized controlled trials (RCTs) and improves the specificity of the search strategy to 29% [106]. This ability to fine tune searches is the result of the National Library of Medicine making improvements to MEDLINE indexing and of initiatives, such as the CONSORT statement, to improve reporting of RCTs.

Another way to decrease the number of nonrelevant citations is to modify the highly sensitive search strategy by excluding publication types that are almost certain not to provide primary data, such as commentaries, editorials, metaanalyses, reviews, or practice guidelines. It has been shown that this approach reduces by 20% the number of nonrelevant citations, without losing any of the relevant trials [107]. The Cochrane strategies have been evaluated carefully and should be consulted (a whole chapter of the Handbook is devoted to searching for randomized trials). There is also a section in the Handbook on searching for nonrandomized studies and in practice this is more difficult. It is also not clear that finding all published studies leads to an unbiased review, since observational studies are not necessarily registered, may not require ethical review, and hence may be completely untraceable. As stated in the Handbook, "Exhaustive searching, which is recommended for randomized trials, may not be justified when reviewing NRS (nonrandomized studies). However, there is no research at present to guide authors about this important issue." It seems likely that searching on methods will be much less fruitful than searching for particular adverse event terms, for example.

Other methods of searching, such as review of the reference sections of retrieved publications found to be relevant, and manual searches of relevant journals, are also recommended.

Establish Inclusion/Exclusion Criteria (Included in "Study selection" in CIOMS X)

A set of rules for including and excluding studies should be defined during the planning stage of the metaanalysis and should be based on the specific hypotheses being tested in the analysis.

One might, for example, wish to limit consideration to randomized studies with more than some minimum number of patients. In a metaanalysis of epidemiologic studies, one might wish to include studies of incident cases only, excluding studies of prevalent cases, assuming that the relationship between exposure and outcome could be different in the two types of study. Practical considerations may, of course, force changes in the inclusion criteria. For example, one might find no randomized studies of a particular new indication for an existing therapeutic agent, thus forcing consideration of nonrandomized studies.

In establishing inclusion/exclusion criteria, one is also necessarily defining the question being addressed by the metaanalysis. If broad inclusion criteria are established, then a broad, and perhaps more generalizable, hypothesis may be tested. The use of broad entry criteria also permits examination of the effects of research design on outcome (e.g., do randomized and nonrandomized studies tend to show different effects of therapy?) or the exploration of subgroup effects. As an example, in a metaanalysis of aspirin administered following myocardial infarction, restriction of the metaanalysis to studies using more than a certain dose of aspirin would not permit an exploratory, cross-study comparison of dose–response effects, which might prove illuminating.

A key point is that exclusion criteria should be based on *a priori* considerations of design of the original studies and completeness of the reports and, specifically, should *not* be based on the results of the studies. To exclude studies solely on the basis of results that contradict the majority of the other studies will clearly introduce bias into the process [12]. While that may seem obvious, the temptation to try to justify such exclusions on a *post hoc* basis may be strong, particularly when a clinically plausible basis for the exclusion can be found. Such exclusions made after having seen the data, and the effect of individual studies on the pooled result, may form the basis for legitimate sensitivity analyses (comparing pooled results with and without that particular study included), but should not be viewed as primary exclusion criteria.

The readers of systematic reviews and metaanalyses often cannot assess whether the exclusion criteria were defined after seeing study results; the registration of systematic reviews protocols will decrease this problem. For example, the Cochrane Collaboration publishes its approved protocols. Cochrane reviews must indicate reasons for deviations from the approved protocol. (Whether the initial question defined in a metaanalysis is motivated, in part, by knowledge of the results of the component studies is a more subtle, and perhaps more important, question.) The realization that reporting of both trials and systematic reviews is often altered by the authors even after a protocol has been recorded shows that the potential for bias is considerable.

Prespecification of the protocol and adherence to it (or at least a clear justification for departures) provides some limited protection against bias. There are registries of protocols for observational studies and one that is specifically for systematic reviews is "PROSPERO" (International **Prospe**ctive **Re**gister **O**f systematic reviews) (www.crd.york.ac.uk/prospero/). PROSPERO is an international database of prospectively registered systematic reviews in health and social care, welfare, public health, education, crime, justice, and international development, where there is a health-related outcome. Key features from the review protocol are recorded and maintained as a permanent record. By August 2018, PROSPERO contained close to 40 000 recorded review protocols. There also is a registry of studies maintained by the European Medicines Agency, under the auspices of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). There is a code of conduct set out by ENCePP, and studies following that code must be registered, but other studies may also register there. It is not

particularly intended for metaanalyses though there are a few included in the register.

Another important note is that studies may often generate more than one published paper. For example, later reports might update analyses previously published, or might report on outcomes not addressed in earlier papers. It is essential, for two reasons, that only one report on the same patients be accepted into the metaanalysis. First, the validity of the statistical methods depends on the assumption that the different studies represent different groups of individuals. Second, the inclusion of a study more than once would assign undue weight to that study in the summary measure. A caution is that it is not always obvious that the same patients have been described in two different publications. Contacting the authors may be of some help in determining if there is duplication, although some authors may perceive the inquiry as questioning their academic integrity. It is also not always obvious what the right choice of report should be for a given study. Certain aspects of the methods may only be reported in earlier publications, which necessitates at least referring to those papers. Methods of analysis may change from paper to paper, or degree of control of confounding, or inclusion or exclusion of certain subpopulations. Thus, there is no general rule we can recommend in such situations, other than trying to exercise good judgment and reporting clearly the reasons for choosing one publication over others. The issue of multiple publications based on the same study has been addressed in more detail by Huston and Moher [108].

Collect the Data

When the relevant studies have been identified and retrieved, the important information regarding study design and outcome needs to be extracted. Typically, data abstraction forms are developed, pilot tested on a few articles, and revised as needed. As in any research, it is necessary to strike a balance between the completeness of the information abstracted and the amount of time needed to extract that information. Careful specification in the protocol for the metaanalysis of the design features and patient characteristics that will be of clinical or academic interest may help avoid over- or undercollecting information. For randomized trials, it is generally advisable, when possible, to collect raw data on outcome measures, such as numbers treated and number of events in each group, rather than derived measures such as odds ratios, which may not be the outcome measures of interest in the metaanalysis or may have been calculated incorrectly by the original authors. For observational studies, in contrast, it will most often be the estimates from each of the studies that are adjusted for confounding that will be of interest. For these studies, the derived measures will be more important to collect, along with information about what confounding factors were included in the adjustment.

Many articles on "how to do a metaanalysis" (e.g., Sacks *et al.* [10], L'Abbe *et al.* [11]), and the PRISMA guidelines) recommend that the metaanalyst assesses the quality of the studies being considered in a metaanalysis. Generally, "quality" is taken to mean freedom from bias, and that terminology has been adopted by PRISMA. Options that have been proposed for incorporating quality in metaanalyses include using a measure of study quality as part of the weight assigned to each study in the analysis, as an exclusion criterion (e.g., excluding studies with quality scores below some arbitrary threshold), or as a stratification factor allowing the separate estimation of effects for good-quality and poor-quality studies [109,110]. Several examples of quality evaluation systems that have been proposed may be of interest [111,112]. Issues related to quality scoring have been discussed more generally by Moher and colleagues [113], and an annotated checklist of quality scoring systems is available [114].

The argument has been made, however, that general scoring systems are arbitrary in their

assignment of weights to particular aspects of study design, and that such systems risk losing information, and can even be misleading [115,116]. Jüni and colleagues, for example, examined studies comparing low molecular weight heparin with standard heparin with respect to prevention of postoperative thrombosis [116]. They used 25 different quality assessment scales to identify high-quality trials. For six scales, the studies identified as being of high quality showed little to no benefit of low molecular weight heparin, while for seven scales, the "high-quality" studies showed a significant advantage of low molecular weight heparin. This apparent contradiction raised questions about the validity of such scales as methods for stratifying studies. One reason why the contradiction arose, the authors argue, is that the quality scores tend to measure a combination of completeness of reporting and factors that might relate to the potential for bias. They recommended, instead, a focus on particular aspects of study design as potential predictors of study outcome, such as whether or not the assessment of outcome is blinded to treatment status.

Thus, in a given metaanalysis, one might wish to examine *specific* aspects of study design that are unique to that clinical or statistical situation [115–118]. For example, Schulz and colleagues found that trials in which the concealment of randomized allocation was inadequate, on average, produced larger estimates of treatment effects compared with trials in which allocation was adequately concealed [119]. This specific finding was not detected when the same authors looked for an overall association between quality score and treatment effect. In the analysis of low molecular weight heparin, Jüni and colleagues found that studies with unmasked outcome assessment showed larger, and presumably biased, benefits of low molecular weight heparin than studies using masked assessment of outcome [116]. Such explorations clearly need to be guided by common sense. As these authors point out, for studies with total mortality as an outcome, masking of outcome assessment would not be expected to impact directly on study findings.

Other authors have suggested essentially similar approaches to that recommended by Jüni and colleagues. For example, Greenland and O'Rourke suggest the use of statistical models to investigate the association between specific design factors and study findings [120]. This approach, known as "response–surface estimation," can be used to derive the predicted outcome for a study with specified (and presumably desirable) characteristics, while at the same time borrowing strength from all the available studies. Once again, caution is needed in performing such analyses with respect to such issues as extrapolation beyond the range of the data. (What if there are *no* studies of sufficient quality on a given dimension included in the model? Is it valid to extrapolate to such studies based on trends observed for lower quality studies?)

The Cochrane Collaboration recommends, via use of the "risk of bias" tool cited above, that authors of systematic reviews assess six domains to determine whether bias in each of the studies included in the analysis is likely to affects their results and hence that of the metaanalysis [18]. These domains are: *selection*, *performance*, *detection*, *attrition*, *reporting*, and *other biases*. Aspects that affect the risk of bias then include the method used to generate the allocation sequence; whether allocation concealment was implemented; whether blinded assessment of outcomes was performed; the degree of completeness of outcome data; whether selective outcome reporting is likely; and "other" dimensions when researchers identify problems that could put the study at a high risk of bias and are not part of the above framework.

Two procedural recommendations have been made regarding the actual techniques for data extraction. One is that studies should be read independently by two readers. The justification

for this comes from metaanalyses in which modest but important interreader variability has been demonstrated [81,82]. A second recommendation is that readers be masked to certain information in studies, such as the identity of the authors and the institutions at which a study was conducted, and masked to the specific treatment assignments or results (when assessing the methods) [60].

While masking has a high degree of intuitive appeal, the effectiveness of masking in avoiding bias has not been demonstrated. Only one randomized trial examined the effect of masking on the results of metaanalyses [121]. This study compared the results of the same metaanalyses performed independently by separate teams of metaanalysts, with one team masked and the other unmasked. The masked and unmasked teams produced nearly identical results on a series of five metaanalyses, lending little support to the need for masking, though it is unclear whether this would be found in all circumstances. Given the propensity for selective analysis and reporting, it seems possible that such selection can take place but equally, it is not clear that masking will prevent this type of bias in practice.

Perform Statistical Analyses

OR, RR or RD – Does It Matter?

There are three summary measures of effect size that can be used in metaanalysis when the outcome of interest is binary (e.g., proportion of subjects with pain relief): relative risk (RR), odds ratio (OR), and risk difference (RD). Although the summary measure used does not affect the statistical significance of the results [122], the choice of effect measure could affect the transferability of results of the metaanalysis into clinical practice. Which summary measure to select depends on the ease of interpretation, the mathematical properties, and the consistency of the results when the particular effect measure is used [123].

Relative risk and RD are easier to interpret than OR. In general, probabilities are more intuitive than odds. When the baseline (untreated) risk is constant across studies, the RD also allows calculation of relevant public health measures (e.g., a number of events prevented or caused by a given treatment). A disadvantage of using RDs in metaanalysis is that, in an empirical study of a large number of metaanalyses, RDs displayed more heterogeneity than ORs, that is, the results from study to study appeared more inconsistent with RDs [124]. Because of this heterogeneity, the extrapolation to a broader population will only be correct at the average baseline risk and extrapolation to other baseline risks will be unreliable. RR and OR are more consistent than RD [122,124] and therefore are preferred from this perspective. There was no difference in heterogeneity in this same sample of metaanalyses, on average, between RR and OR [124].

Odds ratios have better mathematical properties than RRs. For example, switching the roles of the event and nonevent in the analysis is of no consequence for ORs; the new OR is the reciprocal of the original OR (i.e., OR for "benefit" is the reciprocal of the OR for "harm"). In contrast, switching the outcome can make a substantial difference for RR, affecting the treatment effect size and potentially introducing heterogeneity. In a metaanalysis the effect of this reversal cannot be predicted [124].

However, ORs are often incorrectly interpreted as RRs, and this can lead to apparent overestimation of the treatment effect when the outcome is common (when the interpretation is expressed in terms of probabilities, instead of odds). One solution is to discuss the results in terms of RR (or RD) by computing RR (or RD) and confidence intervals from ORs, using the methods described by Localio *et al.* [125].

Choice of Statistical Method

In most situations, the statistical methods for the actual combination of results across studies

are fairly straightforward, although a great deal of literature in recent years has focused on the use of increasingly sophisticated methods. If one is interested in combining odds ratios or other relative measures such as relative risks across studies, for example, some form of weighted average of within-study results is appropriate, and several of these exist [126]. A popular example of this is the Mantel–Haenszel procedure, in which odds ratios are combined across studies with weights proportional to the inverse of the variance of the within-study odds ratio [28,127]. Other approaches include inverse-variance weighted averages of study-specific estimates of covariate-adjusted relative risks and exact stratified odds ratios [126].

Bias in statistical methods is discussed by Tang, who shows that inverse-variance methods may introduce bias in metaanalyses of binary outcomes [128]. Essentially, the problem with those approaches is that the inverse-variance weights depend not only on study size but on the event rates themselves. For example, consider an analysis of 10 trials that all have sample sizes of 500 in both the treated and control groups. Suppose nine studies have event rates of 28% in the treated groups compared with 30% in the control groups. In this same analysis, a single study has event rates of 3% in the treated group versus 1% in controls. For an inverse-variance weighted analysis of risk differences, which are -2% in the nine studies and +2% in the single study, the single study with the low event rates would get 54% of the weight in the metaanalysis, compared with 5.1% of the weight for each of the other nine studies. For an analysis of (log) relative risks, the single study would get 0.4% of the weight, compared with 11.1% of the weight for each of the other nine studies. Appropriate use of weights is also addressed by Chang and colleagues [129].

One basic principle in many analytic approaches is that the comparisons between treated (exposed) and untreated (unexposed) patients are typically made within a study prior to combination across studies. In the combination of randomized trial results, this amounts to preserving the randomization within each study prior to combination. In all the procedures developed for stratified data, "study" plays the role of the stratifying variable. In general, more weight is assigned to large studies than to small studies because of the increased precision of larger studies.

A second basic principle to note is that some of these methods assume that the studies are all estimating a single, common effect, such as a common odds ratio. In other words, the underlying treatment effect (whether beneficial or harmful) that all studies are estimating is assumed to be the same for all studies. Any variability among study results is assumed to be random and is ignored in producing a summary estimate of the treatment effect [130,131]. One may wish to use methods for combining studies that do not make the assumption of a common treatment effect across all studies. These are the so-called "random-effects" models, which allow for the possibility that the underlying true treatment effect, which each study is estimating, may not be the same for all studies, even when examining studies with similar designs, protocols, and patient populations. Hidden or unmeasured sources of among-study variability of results are taken into account by these random-effects models through the incorporation of such variability into the weighting scheme when computing a weighted average summary estimate. Random effects models are described in much greater detail in several papers [132–135].

The practical consequence of the random-effects models is to produce wider confidence intervals than would otherwise be produced by the traditional methods [130,131].This approach is considered particularly useful when there is heterogeneity among study results, and exploratory analyses have failed to uncover any known sources of observed heterogeneity. However, random-effects models should not be viewed as

a panacea for unexplained heterogeneity. One danger is that a summary measure of heterogeneous studies may not really apply to any particular study population or study design; that is, they lose information by averaging over potentially important study and population characteristics [117].

A practical effect of random-effects models, which is only apparent from examining the mathematics involved, is that they tend to assign relatively higher weights to small studies than the traditional methods would assign [130]. This equalization of weights may have unwanted consequences in some circumstances, and can lead to counterintuitive results, with very small studies making contributions to the summary equal to those of very large studies. A thorough discussion of the interpretation and application of fixed-effect versus random-effects models is presented by Hedges and Vevea [136]. Villar and colleagues compared results of fixed-effect and random-effects models on an empirical basis [137]. As expected, in the presence of heterogeneity, they found that the random-effects models gave wider confidence intervals. Interestingly, these random-effects models also showed larger treatment effects than the corresponding fixed-effect models applied to the same data. Explanations for this phenomenon are considered in the section on publication bias.

In many metaanalysis packages (such as Review Manager (RevMan) [138] and comprehensive metaanalysis [139]), the random effects model is implemented with DerSimonian and Laird methodology [1]. However, this methodology is known to be suboptimal in several situations [132,140–145]. Another method that is worth considering is the Hartung–Knapp–Sidik–Jonkman (HKSJ) method for random effects metaanalysis [140–142,145–147]. IntHout *et al*. [147] suggested that the HKSJ method was straightforward and considerably outperforms the standard DerSimonian–Laird method, especially when the number of studies is small, though even with the HKSJ method, extra caution is needed when there are ≤5 studies of very unequal sizes [140–145]. Somewhat in contrast, Bender *et al*. point out that, because heterogeneity is difficult to estimate when there are only a few studies, the HKSJ method accounts appropriately for uncertainty but has very low power [148]. They note, for example, that it is possible for HKSJ to produce a statistically nonsignificant combined result when combining two statistically significant studies with effect estimates going in the same direction. Rover *et al*. proposed a modified Knapp–Hartung (mKH) method that performed well when only a few studies contribute to the metaanalysis and the involved studies' precisions (standard errors) vary [149].

Recently, Stanley and Doucouliagos challenged the two core conventional metaanalysis methods (fixed and random effects) and proposed a weighted least squares method that is neither fixed nor random [150]. They suggested that an unrestricted weighted least squares estimator is superior to conventional random-effects metaanalysis when there is publication (or small-sample) bias and better than a fixed-effect weighted average if there is heterogeneity. It should be noted that, unlike the Peto method, it is unable to deal with zero events in one of the comparison arms.

Another recent publication offers a variety of interpretations of fixed-effect models, going beyond the often-cited limitation that the methods assume a constant treatment effect across all studies [151]. The authors note that some statistical literature considers fixed-effect models appropriate when inferences are to be drawn about the "average" association. They note that Hedges and Vevea [136] offer a rationale for using fixed-effect models even in the presence of heterogeneity. Rice and colleagues go on to recommend exploration of potential sources of variability in treatment effects across studies, using methods described later in this chapter.

Bayesian statistical methods are also being proposed with increasing frequency in the

statistical literature [152–155].These methods can incorporate into the analysis the investigator's prior beliefs about the size of an effect or the factors biasing the observed effects. Bayesian methods are particularly appealing in the metaanalysis setting as they offer the ability to synthesize evidence from multiple sources under a unified framework, to make direct probability statements about any hypotheses, and to handle complex problems [4]. There are a few good practical examples of Bayesian metaanalysis. Askling *et al*. used the Bayesian hierarchical piecewise exponential survival model to investigate the cancer risk for the antitumor necrosis factor (TNF) drug class [156]. The Bayesian model was used to analyze the individual patient-level metadata, taking into account time-dependent covariates and model between study heterogeneity. Kaizar *et al*. used the Bayesian hierarchical model to quantify the risk of suicidality for children who use antidepressants [157]. Ibrahim *et al*. developed a Bayesian metaanalytical method to determine sample sizes for planning a Phase II/III antidiabetic drug development program [158].

Combinability of Results from Diverse Studies: Heterogeneity

The underlying question in any metaanalysis is whether it is clinically and statistically reasonable to estimate an average effect of therapy, either positive or negative. If one errs on the side of being too inclusive, and the studies differ too greatly, there is the possibility that the average effect may not apply to any particular subgroup of patients [159]. Conversely, diversity of designs and results may provide an opportunity to understand the factors that modify the effectiveness (or toxicity) of a drug. Glasziou and Sanders nicely summarize issues related to potential sources of heterogeneity [160]. They highlight an important distinction between artifacts that might be related to either the choice of summary measure or to study design features, and real biological or clinical variation in treatment effect. The former would include issues such as whether relative risk or risk difference is the more appropriate measure of treatment effect, and design issues mentioned above in the context of study quality, such as use of blinding in the evaluation of endpoints within a study. Such features are modifiable aspects of the conduct and analysis of studies. Variation due to clinical factors, in contrast, represents the potential to target therapy to the appropriate patient populations.

With respect to how one should approach the search for sources of heterogeneity, a number of options are available. One might stratify the studies according to patient characteristics or study design features and investigate heterogeneity within and across strata. To the extent that the stratification explains the heterogeneity, the combined results would differ between strata and the heterogeneity within the strata would be reduced compared to the overall result. In addition to stratification, regression methods such as weighted least squares linear regression could be used to explore sources of heterogeneity [3,161–163]. These might be important when various components of study design are correlated with each other, acting as potential confounders. Graphical methods for metaanalysis have also been proposed, that focus on issues related to heterogeneity [164,165].

The quantification of the among-study variability assessment of the degree of variation involves statistical tests. An important word of caution is that statistical tests of heterogeneity suffer from a notorious lack of statistical power [166,167]. Thus, a finding of significant heterogeneity may safely be interpreted as meaning the studies are not all estimating the same parameter. A lack of statistical significance, however, may not mean that heterogeneity is not important in a dataset or that sources of variability should not be explored.

The I^2 statistic seems to have been the most widely adopted approach to statistical

quantification of the among-study variability [168–170]. (Note: In many publications, this is called "between-study variability," based on the popular usage in the statistical literature.) It estimates the proportion of variability in point estimates due to heterogeneity rather than sampling error. The authors recommend I^2 because:

- it focuses attention on the effect of any heterogeneity on the metaanalytic result
- its interpretation is intuitive, that is, the percentage of total variation across studies due to heterogeneity
- it can be accompanied by an uncertainty interval
- it is simple to calculate and can usually be derived from published metaanalyses
- it does not inherently depend on the number of studies in the metaanalysis
- it may be interpreted similarly irrespective of the type of outcome data (e.g., time to event, quantitative, or dichotomous) and choice of effect measure (e.g., OR or hazard ratio).

Several approaches to the statistical modeling of heterogeneity have been proposed. Thompson and Sharp, for example, compared different forms of weighted normal errors regression and random effects logistic regression [171]. Hardy and Thompson reviewed regression methods to investigate heterogeneity [161].

It has been argued that because of the potential for bias in observational epidemiologic studies, exploring heterogeneity should be the main point of metaanalyses of such studies, rather than producing a single summary measure [8,117,172]. (For further information on the metaanalysis of observational studies we refer readers to Section 3.10 of the CIOMS report [4].)

As an example of the type of analysis that could be used to investigate study design issues, Hennessy and colleagues performed a metaanalysis of nonexperimental studies comparing third-generation oral contraceptives (those containing gestodene and desogestrel) to second-generation pills (those containing levonorgestrel) with respect to the risk of venous thromboembolic events [173]. A major issue in these studies has been the possibility of depletion of susceptibles. Specifically, the concern is that users of the newer drugs might tend to be new users of *any* oral contraceptives, whereas users of the older, second-generation drugs would tend to be established users. The risk of venous events tends to be highest for new users, who have events soon after beginning pill use. These susceptible individuals, the argument goes, would be depleted from the ranks of users of second-generation pills, but not from among the third-generation pill users, thereby leaving a more susceptible population of third-generation pill users. The authors found several studies that had performed subgroup analyses of new users in their first year of use. When combined, these subgroups still demonstrated an increased risk from third-generation pills. The power to look within subgroups was only available within the context of the metaanalysis, not within any of the individual studies.

The example just presented was motivated by a specific concern about a hypothesized source of bias in studies. It is sometimes instructive to perform more exploratory analyses of metaanalytic data as well. These may provide valuable insights into the biology of the problem and/or may generate hypotheses for future confirmation.

Analysis of Rare Events

We have mentioned that by combining results of many trials, metaanalysis can address the problems of rare events. However, the analysis of rare events in metaanalysis is still challenging. Many of the methods used to combine data in metaanalysis are based on large sample approximations and therefore may be unsuitable when events are uncommon. In addition, the results could vary substantially depending on the method used to combine the data. Recommendations as to what method to use under which circumstances are

based on studies that have used simulations in which the "truth" is generated by the investigators [174,175]. The results of these studies show that fixed-effect models should be used over random-effect methods [175] and that the inverse-variance-average should be avoided.

When dealing with rare events, many studies may have no events in any of the arms, and relative measures such as relative risk or odds ratios cannot be calculated. If relative measures are used, studies with no events in either treatment arm will be excluded by virtue of the mathematics, not because the metaanalyst chooses to exclude them. However, in these circumstances, risk differences can be estimated. The problem is that risk differences models in the presence of rare events produce biased results and have very limited power [174].

Relative measures in cases when there are no events in *one* arm can be calculated. Many of the methods require "continuity corrections," that is, adding a small value to all cells in a 2×2 table. The Mantel–Haenszel method often uses this approach. Traditionally, 0.5 is added to each of the cells and some statistical packages do this automatically. However, such continuity correction leads to bias in the presence of rare events, and is not necessary, even for the Mantel–Haenszel method [175,176]. Alternative continuity corrections such as the reciprocal of the sample size of the opposite treatment arm, in contrast with the traditional constant continuity correction, produce less biased results [175].

There are methods that do not require using any continuity correction, such as the Peto method and Bayesian methods. The Peto method, also known as the "one-step" model, is a fixed-effect model that focuses on the observed number of events in the experimental intervention and compares it with the expected number of events. Since it uses the expected number of events, it can deal with individual groups in individual trials with no observed events, as long as there is at least one event in at least one of the arms in the trial. The Peto method often produces less biased results provided there is no substantial imbalance between treatment and control group sizes within trials, and provided the treatment effects are not exceptionally large (less than an OR of 5) [175,176].

Bayesian methods can be appropriately applied to rare events metaanalysis. The use of Bayesian hierarchical models can modulate the extremes in the zero-event setting, borrowing information from studies with events to derive posterior inferences for the treatment effect estimates. A practical challenge of Bayesian metaanalysis for rare adverse event data is that noninformative priors may lead to convergence failure due to very sparse data. Weakly informative priors, which put weak restrictions on the size of the treatment effect, may be used to solve this issue [4].

Sensitivity analysis is especially important in the rare AE setting, because results may be sensitive to the choice of statistical methods, scale of measurement, specification of the prior distribution if the Bayesian approach is utilized, and continuity correction factors selected for analyzing zero-events studies, etc. Therefore, when a metaanalysis of rare events is contemplated, a thorough sensitivity analysis pertaining to the above considerations is recommended, and the results of such analyses should be reported so that the readers can assess the robustness of the results [177].

Other Considerations

A number of somewhat specialized statistical issues have been addressed in recent years. These include how to include both parallel and cross-over trials in a single metaanalysis [178–181], the inclusion of trials in which some form of group (e.g., medical practice or hospital) is the unit of randomization (so-called "cluster randomized" trials) in metaanalyses [182], converting odds ratios to effect sizes so that studies with dichotomous outcomes may be combined directly with studies having continuous outcome measures [183], and the analysis of single patient (N-of-1) trials to estimate population

treatment effects and to evaluate individual responses to treatment [184]. Nam and colleagues discuss the analysis of studies with multiple, correlated outcomes [185]. Recently published work of particular interest to epidemiologists includes the analysis of dose–response data from epidemiologic data [186,187], a method for combining disparate designs (case–control, comparative cohort, and uncontrolled cohort studies) [188], and exact methods for case–control and follow-up studies [189].

Formulate Conclusions and Recommendations

As with all research, the conclusions of a metaanalysis should be clearly summarized, with appropriate interpretation of the strengths and weaknesses of the metaanalysis. Authors should clearly state how generalizable the result is and how definitive it is and should outline the areas that need future research. Any hypotheses generated by the metaanalysis should be stated as such, and not as conclusions.

Publication Bias

As discussed earlier, when the primary source of data for a metaanalysis is published data, the possibility needs to be considered that the published studies represent a biased subset of all the studies that have been done. In general, empirical studies have found that it is more likely that studies with statistically significant findings will be published than studies with nonsignificant findings.

A practical technique for determining the potential for publication bias is the "funnel plot," first proposed by Light and Pillemer [190]. The method involves plotting the effect size (e.g., the risk difference) against a measure of study size, such as the sample size or the inverse of the variance of the individual effect sizes. If there is no publication bias, the points should produce a kind of funnel shape, with a scatter of points centered around the true value of the effect size, and with the degree of scatter narrowing as the variances decrease. If publication bias is a problem, the funnel would look as though a bite had been taken out, with very few (if any) points around the point indicating no effect (e.g., odds ratio of 1.0) for studies with large variances. This method requires a sufficient number of studies to permit the visualization of a funnel shape to the data. If the funnel plot does indicate the existence of publication bias, then one or more of the correction methods described below should be considered. In the presence of publication bias, the responsible metaanalyst should also evaluate the ethics of presenting a summary result that is likely to represent an overestimate of the effect in question.

Two examples of funnel plots are given in Figures 36.1 and 36.2. These plots represent studies of psychoeducational programs for surgical patients [190,191]. In the first plot, only the published studies are represented. The funnel appears to have a "bite" taken out of it where the small studies showing no effect of these programs should be. In the second plot, the unpublished studies, including doctoral dissertations, are included, and the former "bite" is now filled with these unpublished studies.

Sterne and Egger provide guidelines for the choice of axes in funnel plots of studies with dichotomous outcomes, recommending that the standard error of the treatment effect (e.g., the standard error of the log odds ratio) be used as the measure of study size and that relative measures (relative risk, as opposed to risk difference) be used as the treatment effect measures [192]. These same authors and a colleague point out that publication bias is only one possible explanation for funnel plot asymmetry, so that the funnel plot should be seen as estimating "small study effects" rather than necessarily publication bias [193]. A similar point is made by Terrin and colleagues [194].

Several mathematical approaches to the problem of publication bias have been proposed. An early method, first described by Rosenthal [195], is the calculation of a "fail-safe N" when the

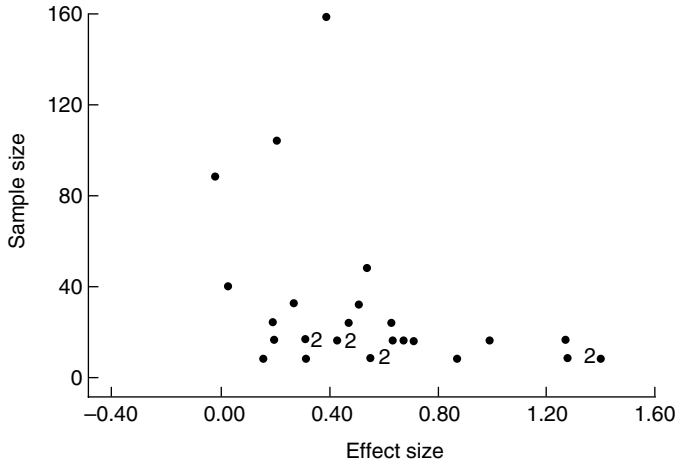


Figure 36.1 Funnel plot for published studies only: analysis of data from Devine and Cook's review of psychoeducational programs for surgical patients [191]. *Source:* Reprinted by permission of the publishers from *Summing Up: The Science of Reviewing Research*, by Richard J. Light and David B. Pillemer, Cambridge, MA: Harvard University Press. ©1984 by the President and Fellows of Harvard College.

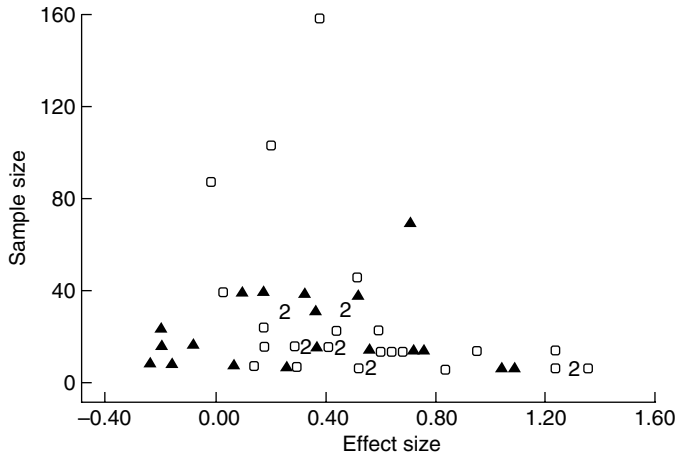


Figure 36.2 Funnel plot for published studies (*open boxes*) and unpublished (*closed triangles*): analysis of data from Devine and Cook's review of psychoeducational programs for surgical patients [191]. *Source:* Reprinted by permission of the publishers from *Summing Up: The Science of Reviewing Research*, by Richard J. Light and David B. Pillemer, Cambridge, MA: Harvard University Press. ©1984 by the President and Fellows of Harvard College.

result of the metaanalysis is a statistically significant rejection of the null hypothesis. This method, in a kind of sensitivity analysis, uses the Z-statistics from the individual studies included in a metaanalysis to calculate the number of *unpublished* studies with a Z-statistic of exactly 0 that would be required to exist, in order for the combined Z-score (published plus unpublished studies) to become nonsignificant. Because this method focuses only on Z-statistics, and ignores the estimation of effects (e.g., odds ratios), it is of limited utility. That is, the fail-safe N approach focuses only on the statistical significance of the combined result and does not help provide an overall estimate of the effect that is "adjusted" for publication bias.

A number of related methods to deal with potential unpublished studies have been developed in recent years. These include other methods for estimating the number of unpublished studies [196,197], formal methods to test for the presence of publication bias [198–200], and methods to adjust summary estimates to account for unpublished studies [136,196,201–203], but several of those methods make some fairly strong assumptions about the specific

mechanism producing the publication bias. A method called "trim-and-fill" has a fair amount of intuitive appeal [204], although it, too, relies on assumptions about the missing studies. It is based on the funnel plot, focusing on the studies that lead to the appearance of funnel plot asymmetry. Under this approach, a mirror image of the studies producing the asymmetry is imputed, using a carefully defined statistical algorithm to determine which studies to mirror, and the impact of adding those mirror image studies to the pooled analysis is assessed.

An additional methodologic caution generated by publication bias relates to the use of random-effects models for combining results. When the results of the studies being analyzed are heterogeneous and a random-effects model is being used to combine those results, one of the properties of the model, described earlier, is to assign relatively higher weights to small studies than would otherwise be assigned by more traditional methods of combining data. If publication bias is a problem in a particular dataset, one consequence implied by the funnel plot is that small studies would tend to show larger effects than large studies. Thus, if publication bias is present, one of the reasons for heterogeneity of study results is that the small studies show systematically larger effects than the large studies. The assignment of higher relative weights to the small studies could, when publication bias is present, lead to a biased summary result.

In fact, this appears to be exactly the situation presented by Poole and Greenland in an examination of studies of water chlorination and cancer [31]. Random-effects summary estimates of the relative risk for various cancers were larger than corresponding fixed-effect summaries. This was apparently due to the assignment of higher relative weights to small studies which, in this case, showed relatively larger effects, that may not be representative of the findings of all small studies. Data presented by Villar and colleagues found a similar phenomenon in studies in perinatal medicine [137].

As noted earlier, one solution to the problem of publication bias is the use of prospective registration of studies at their inception, prior to the availability of results [59]. Others have suggested obtaining unpublished data from the FDA, an approach used by Turner *et al*. [205]. These authors obtained reviews from the FDA for studies of 12 antidepressant agents, conducted a systematic literature search to identify matching publications, and compared the results based on published studies with the results based on the FDA data. They found that among the 74 FDA-registered studies, 31% were not published, and that there was an association between study results and whether or not the paper was published. Of the 38 studies viewed by the FDA as having positive results, 37 were published. Studies viewed by the FDA as having negative or questionable results were, with three exceptions, either not published (22 studies) or published in a way that, in the opinion of the authors of the review, conveyed a positive outcome (11 studies). The analysis restricted to published literature showed that 94% of the trials were positive. In contrast, the analysis of FDA data showed that only 51% were positive.

A further review in looking at other indications for antidepressants found a similar bias in the literature [206]. Although the estimate of effect size was only increased marginally, "Reporting biases led to significant increases in the number of positive findings in the literature." The Open Trials project and the Yale project (YODA) cited earlier are attempts to reduce the bias from using only published literature. There is a tool that allows FDA documents to be retrieved more easily [207].

Going one step further, prospective metaanalyses can be conducted [208–210]. These are metaanalyses that are planned, with complete protocols, including proposed tests of subgroup effects, prior to having knowledge of the results of any of the component studies. More on the topic of prospective metaanalysis is presented later.

Indirect Comparison and Simultaneous Comparison of Treatments Available for Specific Conditions

What healthcare providers, patients, policy makers, and payers often need in order to make informed decisions is to understand how pharmacologic treatments compare to other pharmacologic treatments, even in the absence of direct evidence (head-to-head comparisons). When the treatments of interest have been compared to a common comparator, for example placebo, it is possible to get comparative information via indirect evidence. (See references [211–217] for a series of seven tutorial papers on evidence synthesis for decision making, which includes manuscripts on indirect comparisons – also known as network metaanalysis.)

Indirect evidence involves using data from trials that have compared medication A with medication B, and from trials that have compared medication A with medication C, to draw conclusions about the effect of medication B relative to medication C (Figure 36.3). It is crucial that when an indirect comparison is estimated, the analysis respect the randomization. This means that the analysis must be based on treatment differences within each trial. Pooling the results from the various treatment arms of the clinical trials, by simply collapsing results for that treatment arm across studies, ignores the randomizations and produces biased and overly precise estimates [39]. To correctly assess how medication B compares with medication C, one needs to analyze all the trials that have compared medication A with medication B and calculate (in the case of dichotomous outcome) the appropriate metaanalytic OR and do the same for the trials that have compared medication A with medication C, and then divide these two ORs, that is, OR (B vs C) = OR (A vs B) / OR (A vs C). Section 3.11 of the CIOMS X report also has a high-level summary of the use of indirect evidence [4].

Other advantages of these multiple-treatment comparison techniques are that they can easily deal with trials that have multiple arms and account for the correlation due to multiple arms. In addition to being able to combine direct and indirect evidence, these techniques also permit the assessment of the inconsistency, that is, the disagreement between direct and indirect evidence [218]. These methods can also provide a probabilistic ranking of treatments.

Assumptions

The validity of the indirect comparisons and the extended methodologies we just described depend on meeting assumptions, which are similar to the assumptions of the traditional metaanalysis.

The first assumption is homogeneity. For example, if treatment A in our example is placebo, the results of the placebo-controlled trials that evaluated treatment B should be homogeneous enough to be combined, and the results of the placebo-controlled trials that evaluated

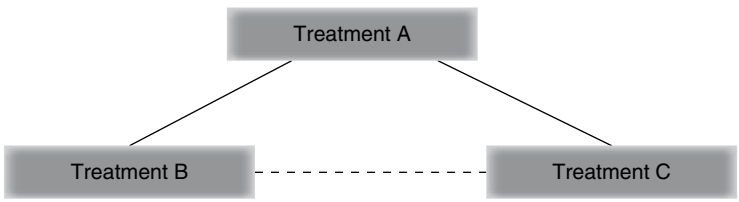


Figure 36.3 Indirect evidence involves using data from trials that have compared medication A with medication B, and from trials that have compared medication A with medication C, to draw conclusions about the effect of medication B relative to medication C (*dotted line*).

treatment C should be homogeneous enough to be combined as well.

The second assumption is similarity. All factors that affect the response to a treatment, effect modifiers, must be similarly distributed across the entire set of trials. This requires that the trials in the network are clinically similar with respect to patient characteristics, settings, follow-up, and outcomes evaluated, and that the trials are methodologically similar, as well. For example, suppose B and C have identical effects, but the size of the treatment effect for both B and C is different in patients with severe disease from that in patients with mild disease. In this situation, variability between studies of B and C with respect to the proportion of patients with severe disease will lead to spurious variability in results between studies of B and studies of C. Similarly, if some trials used enrichment and the others did not, the results are likely to vary across type of study, making questionable the advisability of combining results.

The last assumption to assure validity of the results is consistency (agreement between direct and indirect evidence). It requires that before combining direct and indirect estimates, the consistency of these estimates needs to be checked [39].

Adjusting for Covariates

As we just described, the validity of indirect estimates relies on the balance of factors that affect the response to a treatment in the various treatment arms. When such effect modifiers were measured and reported in the trials, the extended metaanalytic techniques can adjust for possible imbalances of such effect modifiers by incorporating these variables into the statistical mode [219,220]. This is a study-level adjustment for a study-level summary variable (e.g., the proportion of subjects with a particular effect-modifying characteristic), which does not substitute for having access to patient-level characteristics and performing appropriate subgroup analyses, as noted elsewhere in this chapter.

In addition to the study level adjustment for effect modifiers, a more recent approach by Signorovitch and colleagues has been proposed [221,222]. This method, "matching-adjusted indirect comparisons," uses IPD from trials of one treatment to match to baseline summary statistics reported from trials of another treatment; that is, the matching is most often at the level of ensuring the two groups have the same mean and standard deviation. After matching, outcomes can be compared across balanced trial populations by using an approach similar to propensity score weighting. These methods are more frequently being used in health technology assessment submissions. However, more work needs to be done to understand how this new methodology behaves under various scenarios. Philippo *et al*. [223] provided guidance on this methodology that is based on a Technical Support Document prepared for the UK National Institute for Health and Care Excellence Decision Support Unit, available from www.nicedsu.org.uk.

Case Studies of Applications of Metaanalysis

Investigation of Adverse Effects

As mentioned earlier, the investigation of adverse or unwanted effects of existing therapies is an important application of metaanalysis. As discussed in Chapters 1 and 4, adverse events associated with pharmaceutical products are often so uncommon as to be difficult to study. In particular, the usual premarketing randomized studies frequently have too few patients to provide any useful information on the incidence of uncommon adverse events. By the same token, individual studies may have low statistical power to address particular questions. Metaanalysis provides the benefit of increased statistical power to investigate adverse events. In fact, since 1982, the safety evaluation of drugs in the US has included pooled analyses from prospective metaanalysis [224].

The assessment of the cardiovascular safety of rosiglitazone, a medication used to lower blood glucose, provides an excellent example of a situation in which metaanalysis has been both helpful and challenging. The original approval of rosiglitazone was based on its ability to reduce blood glucose levels and glycated hemoglobin levels, and the studies were not powered to determine the effect of this medication on micro- or macrovascular complications of diabetes. To evaluate the effect of rosiglitazone on cardiovascular morbidity and mortality, a metaanalysis was conducted [225].

The authors of this metaanalysis searched published literature, the FDA website, and a clinical trials registry maintained by the drug manufacturer. The authors included RCTs with duration of more than 24 weeks. To combine the data, they used the Peto method. Forty-two trials met the inclusion criteria.

The authors concluded that rosiglitazone increased the risk of myocardial infarction and death from cardiovascular causes. The OR for myocardial infarction was 1.43 (95% CI 1.03–1.98), and the OR for death from cardiovascular causes was 1.64 (95% CI 0.98–2.74).

Not surprisingly, the results of this study generated a great deal of interest. To determine the next course of action, the FDA has reviewed data from observational studies, clinical trials, and the most recently conducted trial called RECORD (**R**osiglitazone **E**valuated for **C**ardiovascular **O**utcomes and **R**egulation of Glycemia in **D**iabetes). The RECORD study was designed to evaluate the cardiovascular safety of rosiglitazone. The FDA presented the results of this review at an Advisory Committee meeting in July 2010. Following this meeting, the FDA announced significant restrictions on the use of rosiglitazone, to patients who cannot control their diabetes on other medications. Under the restricted access program, doctors will have to document their patients' eligibility. Patients will have to review statements describing the cardiovascular safety concerns associated with this drug and acknowledge they understand the risks [226]. In 2013, most of the restrictions on rosiglitazone were removed by the FDA, saying it does "not show an increased risk of heart attack compared to the standard type 2 diabetes medicines metformin and sulfonylurea" based on additional review of the final results of the dedicated CV outcome trial (www.fda.gov/Drugs/DrugSafety/ucm376389.htm).

In 2015, the FDA went further and removed the need for the Risk Evaluation and Mitigation Strategy (REMS) for rosiglitazone-containing medicines. "The REMS is no longer necessary to ensure that the benefits of rosiglitazone medicines outweigh their risks" (www.fda.gov/Drugs/DrugSafety/ucm476466.htm). The license for rosiglitazone remains suspended in the EU in 2018.

This metaanalysis illustrates two of the challenges researchers face when performing metaanalysis: how to deal with rare outcomes and the impact of choosing a method to combine the data.

In this case, the incidence of myocardial infarction in the trials was low. Specifically, observed risks in the rosiglitazone arm ranged from 0 to 1.8% so the authors employed the Peto model to combine the data. As mentioned earlier, this method is recommended in the presence of rare events [174], but it is not recommended when there is substantial imbalance in the number of subjects in the trial arms (unequal treatment allocation), as was the case in this study; some of the studies have an allocation ratio of 4 to 1. When other methods to combine data are used, however, the estimates do not change substantially, but the statistical significance disappears [227]. As we describe in the section on rare events, it is important to assess routinely how robust the results are to the methods used to combine the data and report any discrepancies. We also note in that section (and reinforce here) that a prespecified protocol and careful sensitivity analysis can guard against overreliance on results that are strongly dependent on assumptions and the choice of method for analysis.

New Indications for Existing Therapies

Metaanalysis has also been used to assess the effectiveness of existing therapies for new indications. For example, antidepressants are medications used for the treatment of major depression and other depressive disorders, but they can also reduce pain even in the absence of depression. One of the painful conditions in which antidepressants can be used is fibromyalgia. This is a predominantly female chronic pain condition characterized by widespread pain and tenderness. It can affect up to 10% of women between 55 and 64 years of age [228].

To determine the efficacy of antidepressants in the treatment of fibromyalgia, a metaanalysis of randomized controlled clinical trials was conducted by Hauser and colleagues [229]. The authors searched MEDLINE, PsycINFO, Scopus, the Cochrane Library databases, and reference sections of original studies, metaanalyses, and reviews on antidepressants in fibromyalgia. They included randomized placebo-controlled trials with tricyclic and tetracyclic antidepressants, selective serotonin reuptake inhibitors, serotonin noradrenaline reuptake inhibitors, and monoamine oxidase inhibitors. Two authors independently extracted data. Effects were summarized using standardized mean differences (SMD), analyzed using a random-effects model. The SMD is used to summarize results from studies that used different measurement instruments to assess the same underlying psychiatric construct and are expressed in standard deviation units.

Eighteen randomized controlled trials, with a median duration of eight weeks, involving 1427 participants, were included. The authors found that antidepressants reduced pain intensity (SMD, –0.43; 95% CI –0.55 to –0.30), fatigue (SMD –0.13; 95% CI –0.26 to –0.01), depressed mood (SMD –0.26; 95% CI –0.39 to –0.12), and sleep disturbances (SMD –0.32; 95% CI –0.46 to –0.18). Antidepressants also improved health-related quality of life (SMD –0.31; 95% CI –0.42 to –0.20).

The effect sizes for pain reduction for older antidepressants appeared to be larger than those for the newer drugs. The SMD for tricyclic antidepressants was –1.64 (95% CI –2.57 to –0.71), while the SMD for the newer drugs, such as selective serotonin reuptake inhibitors, was –0.39 (95% CI –0.77 to –0.01) and –0.36 for serotonin and noradrenaline reuptake inhibitors (95% CI –0.46 to –0.25).

This metaanalysis illustrated the utility of metaanalysis for consolidating evidence for new indications for existing therapies. Antidepressants are efficacious for depression and provide short-term relief of fibromyalgia symptoms as well. Serotonin and norepinephrine reuptake inhibitors antidepressants are now approved for the treatment of fibromyalgia. The metaanalysis described here suggests that older antidepressants may be more effective than these drugs, although they also have a different tolerability and safety profile.

Differential Effects Among Subgroups of Patients

Antidepressants labels warn about an increased risk of suicidality in children and adolescents during treatment. To assess this risk in adults, the FDA performed an individual data metaanalysis [230]. Eight industry sponsors of 12 antidepressant products were asked to provide individual data from all completed double blind RCTs of their products, for any indication in adults, with at least 20 participants per arm. Trials limited to known drug responders, such as those using randomized withdrawal designs, were excluded.

Industry sponsors were asked to search their electronic databases for adverse events reported during the double blind phase of treatment, using text strings such as "accident-," "attempt," "burn," "cut," "drown," "gas," "gun," "hang," "hung," "immolat-," "injur-," "jump," "monoxide," "mutilat-," "overdos-," "self damag-," "self harm," "self inflict," "self injur-," shoot," "slash," "suic-," "poison," "asphyxiation," "suffocation," and "firearm." All events identified by this search were considered

possibly related to suicidality, unless they were identified as false positive, that is, events that included any of these text strings but were not related to suicidality, such as "epigastric pain" that would be identified in the search for the text string "gas". The sponsors adjudicated the events. Three individuals blinded to treatment assignment independently rated events. If the three raters were not unanimous in their ratings, a discussion among the raters, led by a fourth rater, was conducted to achieve consensus. In absence of consensus, the event was rated as indeterminate. The FDA staff reviewed the events the sponsors classified as false positives. Events were classified into seven mutually exclusive categories: 1. completed suicide, 2. suicide attempt, 3. preparatory acts towards imminent suicidal behavior, 4. suicidal ideation, 5. Self-injurious behavior, intent unknown, 6. not enough information (fatal), and 7. not enough information (nonfatal). The primary outcome was suicidal ideation or worse (categories 1, 2, 3, or 4). The secondary outcome was suicidal behavior (categories 1, 2, or 3).

Antidepressants were classified *a priori* into five classes: selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors' other modern antidepressants, tricyclic antidepressants, and other antidepressants. Indication was classified into five groups: major depressive disorder, other depressive disorders, other psychiatric disorders, other behavioral disorders, and nonbehavioral disorders.

All the analyses were conditioned on (i.e., stratified by) study. The authors calculated ORs and RDs using conditional logistic regression and other methods, such as exact stratified methods, Mantel–Haenszel, Bayesian, and unconditional and random-effects logistic regression. These multiple methods were used to test the robustness of the findings to the choice of statistical approach. To assess the effect of age on the risk of suicidality, the investigators included age and the interaction of treatment with age as both categorical and continuous variables (in separate models). In addition, the authors performed subgroup analyses based on indication and drug class. To examine heterogeneity of treatment effects across studies, authors added treatment by trial interaction.

The analysis included a total of 99 231 participants in 372 trials, with about 75% of the patients from North America. It is worth noting that most of the studies included were unpublished and, for those that were published, the authors found that they seldom contained information concerning suicidality in the publication.

All the methods to combine the data provided similar results. For participants with nonpsychiatric indications, suicidal behavior and ideation were extremely rare. For those with psychiatric indications, the relative risk of suicidality, associated with treatment, was different for different age groups. The relative risk was higher in participants under 25, neither elevated nor reduced in those aged 25–64 and reduced in those aged 65 and older. For suicidal behavior or ideation, the ORs were 1.62 (95% CI 0.97 to 2.71) for participants aged <25, 0.79 (95% CI 0.64 to 0.98) for those aged 25–64, and 0.37 (95% CI 0.18 to 0.76) for those aged ≥65. For suicidal behavior only, for the same age groupings, the ORs were 2.30 (95% CI 1.04 to 5.09), 0.87 (95% CI 0.58 to 1.29), and 0.06 (95% CI 0.01 to 0.58), respectively. The OR for suicidal behavior or ideation declined 2.6% per year of age (–3.9% to –1.3%), and the OR for suicidal behavior declined 4.6% per year of age (–7.4% to –1.8%). Of note, the increased risk among those <25 years old was larger for suicidal behavior than when ideation was also included, suggesting a stronger association with the more specific definition of the endpoint.

No differences in effect among drugs and drug classes were noted, with the exception of a suggestion of some differences among selective serotonin reuptake inhibitors. Similarly, no difference between older and newer antidepressants was found [230].

This metaanalysis nicely illustrates the amount of effort and regulatory authority involvement necessary to coordinate and gather individual data from a great number of RCTs, involving many drugs and multiple industry sponsors, to assess whether or not a drug *class* increases the risk of a rare but serious outcome and whether or not the increase in risk varies with the characteristic of the subjects exposed. This metaanalysis shows the power of individual data metaanalysis to identify subgroups of patients at higher risk of developing adverse events, and the process of adjudicating adverse events that needs to be followed when the outcome of interest has not been prespecified in the trials or has not been reported in publications.

Saving Time and Resources If You Believe a Metaanalysis

One of the potential benefits of metaanalysis is the ability to shorten the time between a medical research finding and the implementation of regulatory or policy actions or change in clinical practice. This is a concern not only for the development of new drugs, but for the exploration of new indications for existing therapies. As a simple but elegant example of the use of metaanalysis in the approval context, Webber and colleagues reported the use of metaanalysis of ECG data from several clinical pharmacology studies for two drug application submissions [231]. They calculated a pooled estimate for the difference between active doses and placebo on a continuous measure of QT prolongation. This approach allowed the sponsor to avoid having to perform a new safety study to address the question of QT prolongation.

One prominent group has advocated the routine use of what they have termed "cumulative metaanalysis," which is performing a new metaanalysis each time the results of a new clinical trial are published [36,232]. Antman *et al.* applied this technique in combination with a classification scheme of the treatment recommendations for myocardial infarction found in review articles and textbook chapters [36]. They found many discrepancies between the evidence contained in the published randomized trials and the timeliness of the recommendations.

As an example, Antman and colleagues analyzed data from 17 trials of beta-blockers for the prevention of death in the years following a myocardial infarction [36]. In the left-hand side of Figure 36.4, reproduced from their paper, the data are presented as a traditional metaanalysis, with individual study results presented along with the summary odds ratio arbitrarily estimated after 17 trials had been completed. In the right-hand side of Figure 36.4, the same data are presented as a cumulative metaanalysis, with an updated summary estimate calculated after the completion of each new trial. The cumulative metaanalysis clearly shows that the updated pooled estimate became statistically significant in 1977 and has remained so ever since.

Some caution may be advised in interpreting cumulative metaanalyses. The issue of multiple statistical tests, for example, generates concerns about false-positive findings (type I error). In the early papers presenting cumulative metaanalysis, this problem of increased false-positive rates was largely ignored [232–234]. In an empirical study of published literature, Biester and Lange found that only 4% of the cumulative metaanalysis papers even mentioned multiplicity adjustment, and only 2% had made an adjustment [235].

There are several methods proposed to control the false-positive rate. These include sequential approaches [236–238], such as that proposed by Pogue and Yusuf [237], who considered application of conventional group sequential methods [239,240]. These approaches require knowing the number of trials or the sample sizes within the trials. Even when cumulative metaanalyses are planned, those plans can change. More recent discussions on sequential analysis have been published [241–247].

An alternative method was proposed by Lan *et al.*, based on the law of iterated logarithms

(LIL) [248]. That method penalizes the test statistic to account for multiple tests that accounts for estimation of heterogeneity in treatment effects across studies. The initial paper looked at continuous outcomes and was extended by Hu *et al*. for the analysis of relative risk, odds ratios, or risk differences [249]. A limitation of the approach is that there could be a loss of statistical power when the among-study heterogeneity is reasonably small [249] but the extent of the power loss is not always clear. Hu and colleagues recommend using simulation to determine the specific adjustment factor to control the type I error rate but also maintain statistical power.

Another view of cumulative metaanalysis, offered by Lau *et al*. [250], is that the most natural interpretation is a Bayesian framework, in which the existing data form the basis for the prior distribution. When new studies are added, the analysis is updated to generate a posterior distribution, which then becomes the new prior distribution when more data arrive. The concept of multiplicity is handled using prior probabilities for models or hypotheses. Conclusions (usually in the form of "credible intervals") are expressed as probabilistic statements about findings, not as statements about hypotheses.

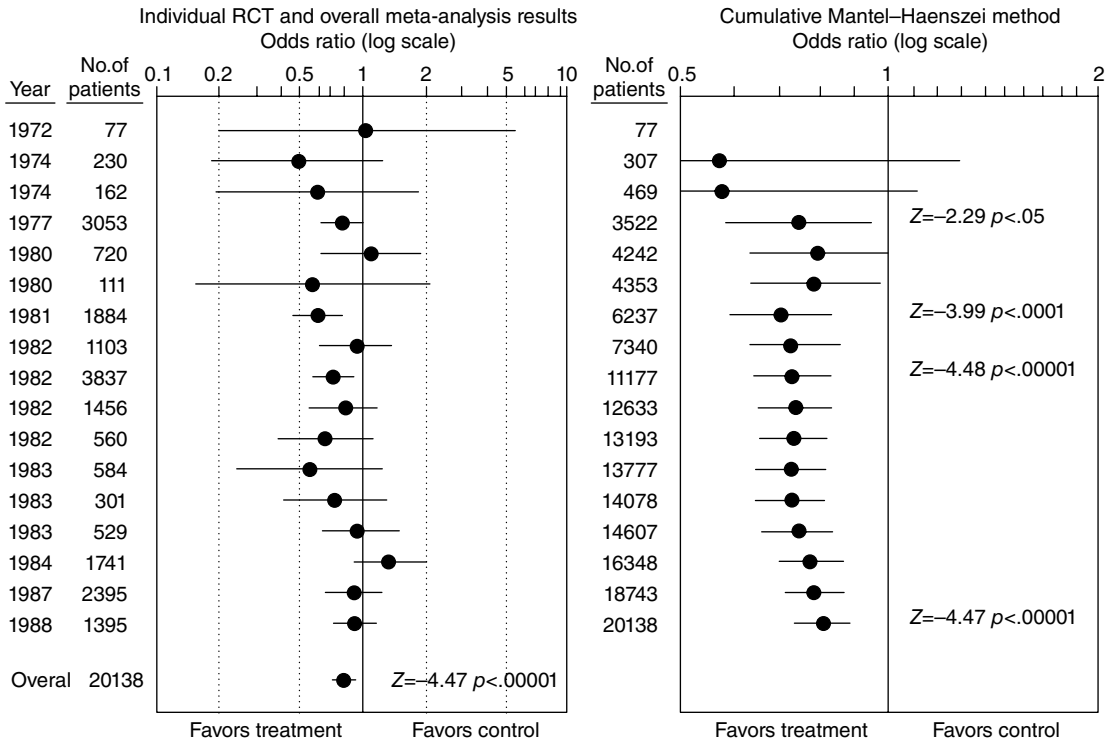


Figure 36.4 Results of 17 RCTs of the effect of oral beta-blockers for secondary prevention of mortality in patients surviving a myocardial infarction presented as two types of metaanalyses. On the left is the traditional one, revealing many trials with nonsignificant results but a highly significant estimate of the pooled results on the bottom of the panel. On the right, the same data are presented as cumulative metaanalyses, illustrating that the updated pooled estimate became statistically significant in 1977 and has remained so up to the present. Note that the scale is changed on the right graph to improve clarity of the confidence intervals. *Source:* Reproduced from Antman *et al*. [36] with permission from the American Medical Association.

There does not seem to be any consensus on a solution to the multiplicity problems generated by cumulative metaanalysis, so some use it mainly as an exploratory tool, providing caveats about the number of analyses performed without a formal correction for multiplicity. This approach is analogous to that for many conventional safety analyses, for which nominal *P* values from hypothesis tests are often provided without adjustment, when there are a limited number of prespecified outcomes.

Another consideration is that estimates of treatment effect may not be stable over time, perhaps due to changing clinical environments. In the beta-blocker example, there is an apparent "drift" of the effect estimate back toward the null in more recent years; that is, treatment appears to be less effective in the most recent studies. Thus, it may be important to reevaluate therapies as other treatment strategies evolve for the same conditions.

A final caution with regard to interpreting cumulative metaanalyses relates to the continuing need for well-designed randomized controlled trials. New indications for existing therapies, for example, are often suggested by nonexperimental studies, including cohort and case–control studies and nonrandomized Phase II clinical trials. The results of these studies are not always confirmed by subsequent, properly designed randomized trials. For example, consider the case of beta-carotene in the prevention of cancer. A series of observational studies (see Ziegler *et al*. [251] for a review) examined the relation between dietary intake of foods rich in beta-carotene and the risk of lung cancer. Overall, they showed a relatively consistent association between diets rich in beta-carotene and reduced risk of lung cancer. Subsequent randomized trials of this specific nutrient as a supplement have failed to confirm a protective effect against lung cancer [252].

For the reasons just outlined, the role of cumulative metaanalysis to demonstrate effectiveness of a therapy in a new indication has not been clarified in actual regulatory settings. Specifically, whether a metaanalysis could be used to support approval of a new indication has not been explicitly addressed. One concern relates to the possibility that the very choice of the question to be investigated may have been influenced by knowledge of the results of the individual studies. Thus, prospective planning of metaanalyses, prior to knowing the results of the component studies, may be useful.

Cumulative Metaanalysis as a Tool to Detect Harm Signals Earlier

Cumulative metaanalysis also could be used as a tool to detect safety signals earlier.

Rofecoxib, a cyclooxygenase-2 inhibitor, was withdrawn from the market in September 2004 because of cardiovascular adverse effects. A cumulative metaanalysis of RCTs was performed to establish whether robust evidence on the adverse effects of rofecoxib was available before its removal. The authors searched bibliographic databases and relevant files of the FDA and included all RCTs in patients with chronic musculoskeletal disorders that compared rofecoxib with other NSAIDs or placebo. Myocardial infarction was the primary outcome [253]. The authors identified 18 randomized controlled trials and found that by the end of 2000 (four years before the withdrawal), the relative risk was 2.30 (95% CI 1.22–4.33), and one year later it was 2.24 (1.24–4.02). The authors found no evidence that the relative risk differed depending on the type of control group (placebo, nonnaproxen NSAID, or naproxen) or trial duration. They concluded that the adverse cardiovascular effects of rofecoxib could have been identified several years earlier, and appropriate action taken.

Cumulative metaanalysis for the evaluation of safety signals brings to light potential methodologic problems that are shared by traditional metaanalysis. First, one might question the validity of pooling of trials that are not clinically homogeneous. For example, the authors

combined the results of trials with dissimilar control arms (placebo, naproxen and nonnaproxen NSAIDs).

Second, the validity of excluding trials that assessed the intervention of interest, but for other indications, can also be questioned. For example, the authors concentrated on trials that evaluated chronic musculoskeletal pain and excluded trials that evaluated Alzheimer's disease. In this case, the inclusion of such a trial would have made the early signal disappear [254]. Clearly, one would not combine trials for different indications to assess efficacy. Although the risks, or relative risks, of potential harms could also vary by indication (population), an approach often used in the regulatory setting is for studies from all indications to be included in at least some safety analyses (perhaps stratified by indication).

Third, one can ask whether efficacy and safety should be evaluated with the same methodologic standards. For efficacy, there are concerns that multiple looks at the data will lead to false-positive results and that *P* values should be adjusted accordingly. When evaluating safety, it could be argued that adjustments to *P* values should not be as large as they are for efficacy analyses (or should not be done at all). A more extensive discussion of the multiplicity issue in safety assessments is presented by Crowe and colleagues in the context of drug development [34]. Additional references can be found in that paper, as well as the discussion earlier on cumulative metaanalysis.

Fourth, it is uncertain whether cumulative metaanalysis (or any metaanalysis of RCTs) can systematically detect harm earlier. Rare adverse events, or the adverse events that occur late after exposure, will likely be absent in RCTs performed during drug development, and therefore cumulative metaanalysis would not always be expected to detect harms earlier.

Ryan and colleagues conducted a metaanalysis of 22 RCTs studying the effects of anti-IL-12/23 therapies [255]. These are antiinflammatory agents used to treat conditions such as psoriasis (the initial indication.) The studies included 10 183 patients. The primary outcome measure was major adverse cardiac events (MACEs). MACE definitions can vary; in this analysis it was defined as a composite of myocardial infarction, cerebrovascular accident, or cardiovascular death during the placebo-controlled portions of the included trials. The authors chose absolute risk differences as their effect measure, using the Mantel–Haenszel fixed-effects method. They found that 10 of 3179 patients receiving anti-IL-12/23 therapies experienced MACEs compared with no events in 1474 patients receiving placebo (Mantel–Haenszel risk difference, 0.012 events/person-year; 95% CI –0.001 to 0.026; $P = 0.12$). (NOTE: in the original paper, the authors use the term "risk difference" but report results in terms of person-time, which would usually require use of rate differences.) They concluded that there was no significant difference in the rate of MACEs associated with anti-IL-12/23 antibodies, but that even the metaanalysis may have been underpowered to identify a significant difference (because there were only 10 events).

In a second metaanalysis, Tzellos and colleagues also studied anti-IL-12/23 biologic agents (ustekinumab and briakinumab) with respect to risk of MACEs, specifically in the setting of treatment of chronic plaque psoriasis [256]. Studies of psoriatic arthritis were excluded, as in the Ryan metaanalysis. These authors used the Peto fixed-effect method to estimate odds ratios. They found a "possible higher risk of MACEs" in patients treated with anti-IL-12/23 antibodies compared with placebo-treated patients (OR 4.23, 95% CI 1.07–16.75, $P = 0.04$).

A lesson from these examples is that, particularly in the setting of rare events, conclusions of the metaanalysis can depend on the inclusion and exclusion criteria for studies and on the choice of statistical methods. Those points have been made throughout this chapter but this is a particularly telling example, because of the potential for impact on the product label.

Indirect Comparisons: Network Metaanalysis and Simultaneous Evaluation of Treatment Therapies for the Same Indication

The efficacy and acceptability of new-generation antidepressants for the treatment of major depression were assessed using multiple treatment metaanalyses. Authors of this metaanalysis included randomized controlled trials that compared 12 new antidepressants and excluded placebo groups where present. Trials were identified in the Cochrane Collaboration Depression, Anxiety, and Neurosis Review Group controlled trials registers, and the authors asked pharmaceutical companies, regulatory agencies, and study investigators to supply information.

Efficacy was evaluated as the proportion of patients who had a reduction of at least 50% from the baseline score on the Hamilton or Montgomery–Åsberg depression rating scales or the proportion of subjects who scored "much" or "very much" improvement on the clinical global impression at eight weeks, or between six and 12 weeks when data at eight weeks were not available. Acceptability of therapy was evaluated as the proportion of patients who terminated the study early for any reason during the first eight weeks of treatment.

The authors calculated the ORs for each of the drugs compared to fluoxetine, using a random-effects model within a Bayesian framework, using Markov chain Monte Carlo methods in WinBUGS [257] (a statistical program). In addition, they estimated the probability that each antidepressant was the most efficacious, or the most acceptable, the second best, the third best, and so on. The Bayesian analysis uses an iterative process to estimate treatment effects. For this analysis, the authors counted the proportion of iterations in which each antidepressant had the highest OR, the second highest, etc., in order to obtain the ranks of treatments in terms of efficacy and acceptability. To assess the consistency between direct and indirect evidence, the authors also calculated the ratio of odds ratios for indirect versus direct evidence.

Overall, 117 trials from 1991 to 2007 with 25 928 individuals assigned to one of the 12 antidepressants were included in the analyses. Overall, there was consistency between direct and indirect evidence. Only three out of 70 comparisons of direct with indirect evidence for efficacy and three out of 63 comparisons for acceptability were found to be inconsistent.

The authors concluded that not all the antidepressants were equally efficacious or equally well tolerated; they provided a matrix that simultaneously compared the 12 antidepressants for efficacy and acceptability and reported the ranking of antidepressants for efficacy or acceptability.

It is not surprising that studies of this nature generate a lot of attention. This study generated many "Letters to the Editor," whose content ranged from congratulations on how well the study helps healthcare providers identify the best treatments to severe criticism. One of the main criticisms was that excluding placebo-controlled data and including only one dose group when multiple doses were evaluated would lead to selection bias that could affect the rank-order of antidepressants. In fact, the ranks were different from those calculated in other studies [258]. Another shortcoming is that publication bias could invalidate the study findings.

Another study compared the results of FDA-registered antidepressant trials with the results from published trials, and found that 95% of the trials in the published literature were "positive" compared to only 51% of FDA-registered studies [205]. Therefore, a metaanalysis that relies primarily on published data, as this study did, will likely overestimate the effect size of treatments.

Food and Drug Administration's Regulatory Role

In recent years, the FDA has used metaanalysis to investigate adverse events associated with the use of certain drugs. The findings from those metaanalyses were used to support a regulatory decision to mandate a labeling change.

As an example, to review the possible association of suicidality events with antiepileptic drugs, the FDA contacted all sponsors of antiepileptic drugs and requested that they submit placebo-controlled trial data from all of their studies. The FDA statistical review of 199 placebo-controlled trials from 11 antiepileptic drugs found that there were 1.9 per 1000 (95% CI 0.6, 3.9) more antiepileptic drug patients than placebo patients who experienced suicidal behavior or ideation compared to the placebo patients (www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM192556.pdf). Based on the findings, the FDA requested the sponsors of antiepileptic drugs, except for those indicated for short-term use, to include new information in the "Warnings and Precautions" section of the product labeling about an increased risk of suicidal thoughts or actions and to develop a Medication Guide to help patients understand this risk.

Not only does metaanalysis sometimes support the decision to change or update the current labeling of approved drugs, it can also provide evidence as to whether or not to keep a drug on the market. A decision may be made either to withdraw the drug completely or to withdraw its use for a particular indication. For example, metaanalysis was used to review the safety of cefepime, which is indicated for treatment of a variety of infections by susceptible strains of microorganisms. Cefepime was suggested to have potentially increased mortality in a study-level metaanalysis published by Yahav *et al*., based on 38 clinical trials [259]. The FDA performed its own metaanalysis on the study level, as well as the patient level, with data from 88 clinical trials. Based on the analysis results, the FDA concluded that cefepime remains an appropriate therapy *for its approved indications*, as neither metaanalysis showed a statistically significant difference in mortality with cefepime.

These examples highlight the point that, while publication bias is often a major concern in conducting a metaanalysis, the FDA has the unique authority to request the sponsors to submit data from all studies performed, regardless of the publication status. An added advantage for this purpose is the FDA's ability to work with patient-level data. As with the antiepileptic drug and the antidepressant cases, the FDA reanalyzed and presented the Nissen–Wolski study-level metaanalysis of rosiglitazone on patient-level data as well (Advisory Committee, July 30, 2007; https://wayback.archive-it.org/7993/20170405051827/www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4308b1-02-fda-backgrounder.pdf). For this metaanalysis, the database for the FDA reanalysis differed on 14 studies compared to the database used in the Nissen–Wolski study; the FDA excluded four open-label trials, six trials that did not include myocardial infarctions or deaths in the analysis, and two long-term trials that were considered not to be suitable for combining with the rest of the short-term, small trials. By updating the database with additional available double-blind, randomized clinical trials, the FDA's reanalysis involved a total of 42 trials that used daily doses of 4 mg or 8 mg of rosiglitazone to treat patients with type 2 diabetes. The FDA's patient-level metaanalysis showed that the overall OR for total ischemic events was 1.4 (95% CI 1.1, 1,8; P = 0.02), and 1.4 for serious ischemic events (95% CI 1.0, 2.1; P = 0.06). These findings were consistent with those of Nissen and Wolski in that about a 40% increase in myocardial ischemia among diabetes patients taking insulin or those using nitrates is observed. However, the FDA reanalysis did not provide sufficient evidence to show an increased risk in the studies comparing rosiglitazone with metformin or a sulfonylurea.

The Future

The examples above have raised several important issues that will need to be addressed in the future. A set of issues not fully addressed above relates to the appropriate approach to evaluating safety during drug development. In particular,

how should the issue of multiplicity be addressed? The SPERT group outlined broad principles and pointed toward potential solutions, including the use of a tiered approach to defining adverse events [34]. During development, there is multiplicity with respect to the enormous number of adverse events that are routinely collected. Literally hundreds of categories are routinely tabulated. If cumulative metaanalyses are updated each time a trial completes during development, the repeated testing (even of events prespecified for formal testing) generates another level of multiplicity. In the safety setting, one would not necessarily want to be as strict in correcting for multiplicity as in the efficacy setting, but at an alpha level of 0.05, the possibility of generating an excessive number of false-positive signals is a real one. Although "compromise" corrections have been proposed, these tend to focus mostly on *P* values, ignoring direct consideration of the magnitude of effects and the clinical importance of the events in question.

We described earlier the current situation with respect to registration of clinical trial protocols and results. There is a wide array of mostly unconnected registries that make it impossible to find all the relevant studies in one place. What would be really useful is a dedicated search engine that, with low false-positive and false-negative rates, would be able to search all these registers to find all the trials with given characteristics, so that those doing a metaanalysis might be able to track unpublished as well as the published trials. It certainly seems there is potential for better ascertainment of *all* the relevant randomized evidence, but it is not clear whether this potential is being met.

When hundreds of categories of events are tabulated, it is likely that most specific events will have been experienced by a very small number of individuals. How broadly or narrowly to define *collections* of events (composite outcomes) becomes a key question in this context. One might wish to err on the side of being inclusive of all types of events that might be related to a drug. Doing so increases the actual counts of events, which can potentially increase statistical power. Conversely, choosing a narrower definition risks being too granular and losing statistical power by reducing the counts, but may also eliminate "noise" (events that are clinically less important or that may simply be associated with the underlying indication). Work to date suggests that more targeted definitions can sometimes lead to stronger signals (larger relative risks) and may actually make it more likely that signals will be detected [260,261].

The question of how to respond, from a sponsor or regulatory perspective, in the presence of heterogeneous results is also an open one. When there is little or no heterogeneity of results among trials, one might be willing to accept metaanalytic evidence as helping to establish effectiveness or harm. It is less obvious what to do with the results of a metaanalysis when there is substantial heterogeneity. If the heterogeneity is adequately explained in the analysis in terms of subgroup effects or trial quality, metaanalysis might still be an acceptable part of demonstrating effectiveness or harm, but such a conclusion might be conditional on the type of patient or other factors. How should results be interpreted when some trials show harm and others show no effect of a drug (relative risks or risk differences close to the null)? Is this an indication that treatment is harmful in some but not all situations? Does such a situation simply reflect random variability? The threshold for action in the face of heterogeneity of findings may well be different for safety endpoints than for efficacy endpoints, but work is needed to establish transparent criteria by which to evaluate such situations.

Earlier in this chapter we discussed the principles behind indirect comparisons. As the focus of policy and clinical decisions moves in the direction of comparative effectiveness (see Chapter 26), which also includes comparative safety, there are serious questions about how to define research agendas. In principle, one might

wish to make direct comparisons across all drugs (or therapies) for a given indication. Who will fund such studies, which will need to be large, is not at all clear. The principles defining validity of indirect comparisons have been described. Work is needed, however, to explore in practice the conditions under which indirect comparisons, or mixed treatment comparisons, may be both valid and useful. Are there particular types of questions that can be evaluated using these alternative approaches? One study showed that indirect comparisons often, but not always, agree with direct comparisons [262]. How and when to incorporate studies that are not head-to-head comparisons needs further empirical study.

The inclusion of nonexperimental observational studies in metaanalyses, particularly of serious but uncommon adverse events, will almost certainly be a necessity. To the extent that clinical trials performed in support of new drug approvals tend to include populations that are different from the population in which the drug will be used after approval, safety assessments done during development will need to be supplemented with studies done in actual clinical practice. Sample sizes during development also tend to be limited, making it necessary to study large populations to evaluate risks of uncommon but serious adverse events.

In the US, the FDA has established a network of observational databases, known as the Sentinel Network, aimed at exactly this type of assessment of drug safety in clinical practice (described in detail in Chapter 25). The legislative mandate was to provide access to claims or electronic medical records data from 100 million individuals by 2012 [263,264]. Current information regarding Sentinel can be accessed at the Sentinel website (www.fda.gov/Safety/FDAsSentinelInitiative/default.htm) [265]. Sentinel posts results of analyses conducted and provides open access to analytical tools used in those analyses.

Sentinel uses what is known as a distributed network (see Chapter 25). That is, providers of data house the data and provide analytical results, at the aggregate level only, to a central group that evaluates the appropriateness of combining results across data sources. A distributed network allows the data providers, who are most familiar with the idiosyncrasies of their respective databases, to be the ones manipulating the raw data. This approach also avoids issues related to privacy, as only the aggregate-level results are made public. Making such a distributed approach work efficiently and effectively requires the use of a common data model, that is, shared definitions of variables related to drug exposure and outcomes, across all data sources.

In a related effort, a public–private partnership, known as the Observational Medical Outcomes Partnership (OMOP), was funded by the pharmaceutical industry and included representatives from industry, the FDA, and academic institutions. OMOP investigators conducted methodologic research to determine which approaches to analysis of such observational data provide the "best" (least biased, most consistent, most precise) results (see Chapter 27). OMOP also adopted a common data model, which differs from the Sentinel common data model.

OMOP no longer exists but gave rise to two related efforts: Innovation in Medical Evidence Development and Surveillance (IMEDS: http://reaganudall.org/innovation-medical-evidence-development-and-surveillance) and the Observational Health Data Sciences and Informatics (OHDSI: www.ohdsi.org/) been established under the auspices of the Reagan–Udall Foundation. IMEDS is a public–private partnership that now provides access for private-sector organizations, such as the pharmaceutical industry, academic institutions, and nonprofit organizations to a system based on Sentinel. Selected Sentinel data partners and the Harvard Pilgrim Healthcare Institute, which functions as the analytic or coordinating

center, with governance by IMEDS, facilitate the analyses of medical product safety evaluations. The Reagan–Udall Foundation was established by the US Congress to advance regulatory science, for the purpose of helping the FDA advance its mission.

The OHDSI (pronounced Odyssey) program is a multi-stakeholder collaborative established to bring out the value of health data through large-scale analytics. All of its solutions are open-source. OHDSI has established a global network of researchers and observational health databases with a coordinating center housed at Columbia University. All the data sources available to OHDSI are part of a distributed data network and use the OMOP common data model.

Other distributed data networks have been established outside the US for similar purposes. For example, the Drug Safety and Effectiveness Network (DSEN) in Canada created a pan-Canadian collaboration of researchers, the Canadian Network for Observational Drug Effect Studies (CNODES: www.ices.on.ca/Research/Research-programs/Chronic-Disease-and-Pharmacotherapy/CNODES), to facilitate the study of specific drug safety and effectiveness questions using multiple (generally provincial) healthcare databases. The overarching aim of CNODES is to use collaborative approaches to provide rapid, population-based answers to questions about drug safety and effectiveness. The Asian Pharmacoepidemiology Network (ASPEN: http://aspennet.asia/) has also been created as a multinational research network. Similar to the other networks, it was formed to support the conduct of drug safety research and help identify emerging safety issues among the Asian countries. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP; www.ema.europa.eu/ema/index.jsp?curl=pages/partners_and_networks/general/general_content_000229.jsp&mid=WC0b01ac05801df747) was established in 2007 and also aims to strengthen the ongoing evaluation of benefits and risks of medicines, principally by facilitating the conduct of multicenter, independent (of the sponsors) post-authorization studies, almost all of which are observational research.

In conclusion, while there are no easy answers to many of the questions presented in this chapter, it is clear that metaanalysis will play an increasingly important role in the formulation of treatment and policy recommendations. Thus, the qualities of the metaanalyses performed, and of the included studies, are of the utmost importance and need to be reviewed by the scientific community in an open, published forum. If they are carefully interpreted in view of their strengths and weaknesses, metaanalyses should prove to be extremely helpful in pharmacoepidemiologic research.

References

1 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**(3): 177–88.

2 Last JM. *A Dictionary of Epidemiology*, 2nd edn. New York: Oxford University Press, 1988.

3 Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose–response data. *Epidemiology* 1993; **4**(3): 218–28.

4 CIOMS Working Group X. *Evidence Synthesis and Meta-analysis for Drug Safety. Report of the CIOMS Working Group X*. Geneva: Council for International Organizations of Medical Sciences (CIOMS), 2016.

5 Golub RM, Fontanarosa PB. Researchers, readers, and reporting guidelines: writing between the lines. *JAMA* 2015; **313**(16): 1625–6.

6 Feinstein AR. Para-analysis, faute de mieux, and the perils of riding on a data barge. *J Clin Epidemiol* 1989; *42*(10): 929–35.

7 Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997; **50**(3): 2239.

8 Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**: 140–4.

9 Buffler PA. The evaluation of negative epidemiologic studies: the importance of all available evidence in risk characterization. *Regul Toxicol Pharmacol* 1989; **9**(1): 34–43.

10 Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; **316**(8): 450–5.

11 L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; **107**(2): 224–33.

12 Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. *Am J Epidemiol* 1995; **142**(4): 371–82.

13 Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995; **142**(4): 3837.

14 Bero LA, Rennie D. The Cochrane Collaboration – preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; **274**: 1935–8.

15 Doyle J, Waters E, Noonan E, *et al*. Working with The Campbell Collaboration to produce reviews of relevance to public health across the education, justice and social welfare sectors. *J Public Health* 2010; **32**(1): 138–40.

16 Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014; **4**: MR000034.

17 Loke YK, Price D, Herxheimer A, Group CAEM. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007; **7**: 32.

18 Higgins JP, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Oxford: Cochrane Collaboration, 2011.

19 Trialists'Collaboration A. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002; **324**(7329): 71–86.

20 Canner PL. An overview of six clinical trials of aspirin in coronary heart disease. *Stat Med* 1987; **6**(3): 255–67.

21 Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988; **260**(5): 652–6.

22 Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol* 1990; **132**(4): 612–28.

23 Morgenstern H, Glazer WM, Niedzwiecki D, Nourjah P. The impact of neuroleptic medication on tardive dyskinesia: a meta-analysis of published studies. *Am J Pub Health* 1987; **77**(6): 717–24.

24 Larsson SC, Orsini N, Wolk A. Vitamin B6 and risk of colorectal cancer: a meta-analysis of prospective studies. *JAMA* 2010; **303**(11): 1077–83.

25 Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology – a proposal for reporting. *JAMA* 2000; **283**(15): 2008–12.

26 Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. *JAMA* 1999; **281**(9): 830–4.

27 Greenland S, Longnecker MP. Methods for trend estimation from summarized dose–response data, with applications to meta-analysis. *Am J Epidemiol* 1992; **135**(11): 1301–9.

28 Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med* 1990; **9**(3): 247–52.

29 Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies

of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991; **44**(2): 127–39.

30 O'Neill RT, Anello C. Does research synthesis have a place in drug regulatory policy? Synopsis of issues: assessment of efficacy and drug approval. *Clin Res Regul Aff* 1996; **13**: 23–9.

31 Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol* 1999; **150**(5): 469–75.

32 Yakoob MY, Bateman BT, Ho E, *et al*. The risk of congenital malformations associated with exposure to beta-blockers early in pregnancy: a meta-analysis. *Hypertension* 2013; **62**(2): 375–81.

33 Crowe B, Chuang-Stein C, Lettis S, Brueckner A. Reporting adverse drug reactions in product labels. *Therapeut Innov Regul Sci* 2016; **50**(4): 455–63.

34 Crowe BJ, Xia HA, Berlin JA, *et al*. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials* 2009; **6**(5): 430–40.

35 Lis Y, Roberts MH, Kamble S, Guo JJ, Raisch DW. Comparisons of Food and Drug Administration and European Medicines Agency risk management implementation for recent pharmaceutical approvals: report of the International Society for Pharmacoeconomics and outcomes research risk benefit management working group. *Value Health* 2012; **15**(8): 1108–18.

36 Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; **268**(2): 240–8.

37 Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. *Int J Technol Asses. Health Care* 2008; **24**(2): 170–7.

38 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; **23**: 3105–24.

39 Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; **338**: b1147.

40 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; **151**(4): W65–W94.

41 Higgins JP, Altman DG, Gotzsche PC, *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; **343**: d5928.

42 Whiting P, Savović J, Higgins JPT, *et al*. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016; **69**: 225–34.

43 Sterne JA, Hernan MA, Reeves BC, *et al*. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016; **355**: i4919.

44 Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open* 2018; **8**(3): e019703.

45 Sabatine MS, Wiviott SD, Im K, Murphy SA, Giugliano RP. Efficacy and safety of further lowering of low-density lipoprotein cholesterol in patients starting with very low levels: a meta-analysis. *JAMA Cardiol* 2018; **3**(9): 823–8.

46 Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989; **81**(2): 107–15.

47 Andrews TC, Reimold SC, Berlin JA, Antman EM. Prevention of supraventricular arrhythmias after coronary artery bypass surgery. A meta-analysis of randomized control trials. *Circulation* 1991; **84**(5 Suppl): III 236–44.

48 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**(8746): 867–72.

49 Moher D, Berlin J. Improving the reporting of randomised controlled trials. In: Maynard A, Chalmers I, eds. *Non-Random Reflections on Health Services Research*. London: BMJ Publishing Group, 1997, pp. 250–71.

50 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992; **267**(3): 374–8.

51 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; **263**(10): 1385–9.

52 Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993: Doc. No. 50.

53 Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990; **263**(10): 1405–8.

54 Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990; **263**(10): 1392–5.

55 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; **315**(7109): 640–5.

56 Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; **279**(4): 281–6.

57 Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989; **84**: 381–92.

58 Sutton AJ, Abrams KR, Jones DR. Generalized synthesis of evidence and the threat of dissemination bias: the example of electronic fetal heart rate monitoring (EFM). *J Clin Epidemiol* 2002; **55**: 1013–24.

59 De Angelis C, Drazen JM, Frizelle FA, *et al*. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med* 2004; **351**(12): 1250–1.

60 Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. II: Replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987; **6**(7): 733–44.

61 Group EBCTC. *Treatment of Early Breast Cancer. Vol I. Worldwide Evidence*. New York: Oxford University Press, 1990.

62 Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ* 2005; **331**(7507): 19.

63 Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009; **1**: MR000006.

64 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000; **320**(7249): 1574–7.

65 Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989; **84**(2): 374–80.

66 Gotzsche PC. Reference bias in reports of drug trials. *BMJ Clin Res Ed* 1987; **295**(6599): 654–6.

67 Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; **350**(9074): 326–9.

68 Moher D, Fortin P, Jadad AR, *et al*. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; **347**(8998): 363–6.

69 Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986; **1**(3): 155–8.

70 Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980; **280**(6217): 833–6.

71 Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996; **124**(5): 485–9.

72 O'Connor AB. The need for improved access to FDA reviews. *JAMA* 2009; **302**(2): 191–3.

73 Ladanie A, Ewald H, Kasenda B, Hemkens LG. How to use FDA drug approval documents for evidence syntheses. *BMJ* 2018; **362**: k2815.

74 Hutton J, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *J Roy Stat Soc C: Appl Stat* 2000; **49**(3): 359–70.

75 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by The Lancet. *Lancet* 2008; **372**(9634): 201.

76 Dwan K, Altman DG, Clarke M, *et al*. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014; **11**(6): e1001666.

77 Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. *PLoS One* 2010; **5**(3): e9810.

78 Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014; **349**: g6501.

79 Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med* 2008; **5**(11): e217.

80 Amarilyo G, Furst DE, Woo JMP, *et al*. Agreements and discrepancies between FDA reports and journal papers on biologic agents approved for rheumatoid arthritis: a meta-research project. *PLoS One* 2016; **11**(1): e0147556.

81 Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991; **10**(6): 971–9; discussion 979–80.

82 Chalmers TC, Berrier J, Hewitt P, *et al*. Meta-analysis of randomized controlled trials as a method of estimating rare complications of non-steroidal anti-inflammatory drug therapy. *Aliment Pharmacol Ther* 1988; **2**(Suppl): 9–26.

83 DerSimonian R. Parenteral nutrition with branched-chain amino acids in hepatic encephalopathy: meta analysis. *Hepatology* 1990; **11**(6): 1083–4.

84 Naylor CD, O'Rourke K, Detsky AS, Baker JP. Parenteral nutrition with branched-chain amino acids in hepatic encephalopathy. A meta-analysis. *Gastroenterology* 1989; **97**(4): 1033–42.

85 Erikkson LS, Conn HO. Branched-chain amino acids in hepatic encephalopathy. *Gastroenterology* 1990; **99**(2): 604–7.

86 Jelicic Kadic A, Vucic K, Dosenovic S, Sapunar D, Puljak L. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. *J Clin Epidemiol* 2016; **74**: 119–23.

87 Stewart LA, Clarke M, Rovers M, *et al*. Preferred Reporting Items for Systematic Review and Meta–Analyses of individual participant data: the PRISMA–IPD Statement. *JAMA* 2015; **313**(16): 1657–65.

88 Early Breast Cancer Trialists' Collaborative Group, Darby S, McGale P, *et al*. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet* 2011; **378**(9804): 1707–16.

89 Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001; **358**(9291): 1389–99.

90 Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of

treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002; **21**(3): 371–87.

91 Berlin JA, Morris S, Rockhold F, Askie L, Ghersi D, Waldstreicher J. Bumps and bridges on the road to responsible sharing of clinical trial data. *Clin Trials* 2014; **11**(1): 7–12.

92 Goldacre B. Commentary on Berlin *et al*. *Clin Trials* 2014; **11**(1): 15–18.

93 Platt R, Lieu T. Data enclaves for sharing information derived from clinical and administrative data. *JAMA* 2018; **320**: 753–4.

94 Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J Med Ethics* 2015; **41**(5): 404–9.

95 Thacker SB. Meta-analysis. A quantitative approach to research integration. *JAMA* 1988; **259**(11): 1685–9.

96 Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009; **62**(10): 1006–12.

97 Zorzela L, Loke YK, Ioannidis JP, *et al*. PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ* 2016; **352**: i157.

98 Hutton B, Salanti G, Caldwell DM, *et al*. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015; **162**(11): 777–84.

99 Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS One* 2013; **8**(12): e83138.

100 Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. *Syst Rev* 2017; **6**(1): 263.

101 Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2*. Oxford: Cochrane Collaboration, 2009.

102 Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007; **7**: 16.

103 Bernstein F. The retrieval of randomized clinical trials in liver diseases from the medical literature: manual versus MEDLARS searches. *Control Clin Trials* 1988; **9**(1): 23–31.

104 Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Control Clin Trials* 1985; **6**(4): 306–17.

105 Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. *Control Clin Trials* 1985; **6**(4): 271–9.

106 Eisinga A, Siegfried N, Clarke M. The sensitivity and precision of search terms in Phases I, II and III of the Cochrane Highly Sensitive Search Strategy for identifying reports of randomized trials in medline in a specific area of health care – HIV/AIDS prevention and treatment interventions. *Health Info Libr J* 2007; **24**(2): 103–9.

107 Biondi-Zoccai GG, Agostoni P, Abbate A, Testa L, Burzotta F. A simple hint to improve Robinson and Dickersin's highly sensitive PubMed search strategy for controlled clinical trials. *Int J Epidemiol* 2005; **34**(1): 224–5.

108 Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996; **347**(9007): 1024–6.

109 Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; **45**(3): 255–65.

110 Berard A, Bravo G. Combining studies using effect sizes and quality scores: application to bone loss in postmenopausal women. *J Clin Epidemiol* 1998; **51**: 801–7.

111 Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Assoc J* 1964; **90**: 565–9.

112 Verhagen AP, de Vet HCW, de Bie RA, *et al*. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; **51**: 1235–41.

113 Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care* 1996; **12**(2): 195–208.

114 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995; **16**(1): 62–73.

115 Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; **140**: 300–1.

116 Juni P, Witschi A, Block R, Egger M. The hazards of scoring the quality of clinical trials: lessons for meta-analysis. *JAMA* 1999; **282**(11): 1054–60.

117 Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; **140**(3): 290–6.

118 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**(5): 408–12.

119 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002; **359**(9306): 614–18.

120 Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001; **2**: 463–71.

121 Berlin JA, Miles CG, Cirigliano MD, *et al*. Does blinding of readers affect the results of metaanalyses? Results of a randomized trial. *Online J Curr Clin Trials* 1997: Doc. No. 205.

122 Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000; **19**(13): 1707–28.

123 Altman DG, Deeks JJ. Meta-analysis, Simpson's paradox, and the number needed to treat. *BMC Med Res Methodol* 2002; **2**: 3.

124 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002; **21**(11): 1575–600.

125 Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *J Clin Epidemiol* 2007; **60**: 874–82.

126 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. New York: Van Nostrand Reinhold, 1982.

127 Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; **1**: 43–6.

128 Tang J-L. Weighting bias in meta-analysis of binary outcomes. *J Clin Epidemiol* 2000; **53**: 1130–6.

129 Chang BH, Waternaux C, Lipsitz S. Meta-analysis of binary data: which within study variance estimate to use? *Stat Med* 2001; **20**(13): 1947–56.

130 Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989; **8**(2): 141–51.

131 Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res* 1993; **2**(2): 121–45.

132 Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics* 1999; **55**(3): 732–7.

133 Aitkin M. Meta-analysis by random effect modelling in generalized linear models. *Stat Med* 1999; **18**(17–18): 2343–51.

134 Frost C, Clarke R, Beacon H. Use of hierarchical models for meta-analysis: experience in the metabolic ward studies of diet and blood cholesterol. *Stat Med* 1999; **18**: 1657–76.

135 Senn SJ. The many modes of META. *Drug Inf J* 2000; **34**: 535–49.

136 Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods* 1998; **3**: 486–504.

137 Villar J, Mackey ME, Carroli G, Donner A. Meta-analyses in systematic reviews of randomized controlled trials in perinatal medicine: comparison of fixed and random effects models. *Stat Med* 2001; **20**: 3635–47.

138 Review Manager (RevMan). Version 5.3: Copenhagen: Nordic Cochrane Centre, 2014.

139 Comprehensive Meta-analysis. Version 2: Englewood, NJ: Biostat, 2005.

140 Hartung J. An alternative method for meta-analysis. *Biometric J* 1999; **41**(8): 901–16.

141 Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med* 2001; **20**(12): 1771–82.

142 Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med* 2001; **20**(24): 3875–89.

143 Hartung J, Makambi KH. Reducing the number of unjustified significant results in meta-analysis. *Commun Stat Simul Comput* 2003; **32**(4): 1179–90.

144 Makambi KH. The effect of the heterogeneity variance estimator on some tests of treatment efficacy. *J Biopharm Stat* 2004; **14**(2): 439–49.

145 Sidik K, Jonkman JN. Robust variance estimation for random effects meta-analysis. *Comput Stat Data Anal* 2006; **50**(12): 3681–701.

146 Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med* 2002; **21**(21): 3153–9.

147 IntHout J, Ioannidis JP, Borm GF. The Hartung–Knapp–Sidik–Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian–Laird method. *BMC Med Res Methodol* 2014; **14**: 25.

148 Bender R, Friede T, Koch A, *et al*. Methods for evidence synthesis in the case of very few studies. *Res Synth Methods* 2018; **9**(3): 382–92.

149 Rover C, Knapp G, Friede T. Hartung–Knapp–Sidik–Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Med Res Methodol* 2015; **15**: 99.

150 Stanley TD, Doucouliagos H. Neither fixed nor random: weighted least squares meta-analysis. *Stat Med* 2015; **34**(13): 2116–27.

151 Rice K, Higgins J, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J Roy Stat Soc A* 2018; **181**(1): 205–27.

152 Carlin JB. Meta-analysis for 2 x 2 tables: a Bayesian approach. *Stat Med* 1992; **11**(2): 141–58.

153 Larose DT, Dey DK. Grouped random effects models for Bayesian meta-analysis. *Stat Med* 1997; **16**(16): 1817–29.

154 Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995; **14**(24): 2685–99.

155 Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. *Med Decis Mak* 1990; **10**(1): 15–23.

156 Askling J, Fahrbach K, Nordstrom B, Ross S, Schmid CH, Symmons D. Cancer risk with tumor necrosis factor alpha (TNF) inhibitors: meta-analysis of randomized controlled trials of adalimumab, etanercept, and infliximab using patient level data.

Pharmacoepidemiol Drug Saf 2011; **20**(2): 119–30.

157 Kaizar EE, Greenhouse JB, Seltman H, Kelleher K. Do antidepressants cause suicidality in children? *A Bayesian meta-analysis. Clin Trials* 2006; **3**(2): 73–98.

158 Ibrahim JG, Chen MH, Xia HA, Liu T. Bayesian meta-experimental design: evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. *Biometrics* 2012; **68**(2): 578–86.

159 Simon R. Overviews of randomized clinical trials. *Cancer Treatment Rep* 1987; **71**(1): 3–5.

160 Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; **21**: 1503–11.

161 Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med* 1998; **17**(8): 841–56.

162 Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. *Online J Curr Clin Trials* 1994; Doc No 134.

163 Vanhonacker WR. Meta-analysis and response surface extrapolation: a least squares approach. *Am Stat* 1996; **50**: 294–9.

164 Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. *Am J Pub Health* 1988; **78**(8): 961–6.

165 Song F. Exploring heterogeneity in meta-analysis: is the L'Abbe plot useful? *J Clin Epidemiol* 1999; **52**(8): 725–30.

166 Fleiss JL. Analysis of data from multiclinic trials. *Control Clin Trials* 1986; **7**(4): 267–75.

167 Takkouche B, Cadarso-Suarez C, Spiegelman D. Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *Am J Epidemiol* 1999; **150**(2): 206–15.

168 Higgins JP, Spiegelhalter DJ. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *Int J Epidemiol* 2002; **31**(1): 96–104.

169 Higgins J, Thompson SG, Deeks J, Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003; **327**(September 6): 557–60.

170 Higgins J, Thompson SG, Deeks J, Altman DG. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* 2002; **7**(1): 51–61.

171 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18**: 2693–708.

172 Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? *BMJ* 1997; **315**(7122): 1610–14.

173 Hennessy S, Berlin JA, Kinman JL, Margolis DJ, Marcus SM, Strom BL. Risk of venous thromboembolism from oral contraceptives containing gestodene and desogestrel versus levonorgestrel: a meta-analysis and formal sensitivity analysis. *Contraception* 2001; **64**(2): 125–33.

174 Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007; **26**(1): 53–77.

175 Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004; **23**(9): 1351–75.

176 Sweeting MJ, Sutton AJ, Lambert PC. Correction. *Stat Med* 2006; **25**: 2700.

177 Dahabreh IJ, Economopoulos K. Meta-analysis of rare events: an update and sensitivity analysis of cardiovascular events in randomized trials of rosiglitazone. *Clin Trials* 2008; **5**(2): 116–20.

178 Curtin F, Altman DG, Elbourne D. Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. *Stat Med* 2002; **21**: 2131–44.

179 Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes. *Stat Med* 2002; **21**: 2145–59.

180 Elbourne D, Altman DG, Higgins J, Curtin F, Worthington H, Vail A. Meta-analysis involving cross-over trials: methodological issues. *Int J Epidemiol* 2002; **31**: 140–9.

181 Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. III: The issue of carry-over. *Stat Med* 2002; **21**: 2161–73.

182 Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Stat Med* 2002; **21**: 2971–80.

183 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000; **19**: 3127–31.

184 Zucker D, Schmid CH, McIntosh MW, d'Agostino RB, Selker HP, Lau J. Combining single patient (n-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *J Clin Epidemiol* 1997; **50**(4): 401–10.

185 Nam I-S, Mengersen KL, Garthwaite P. Multivariate meta-analysis. *Stat Med* 2003; **22**: 2309–33.

186 Cook R, Brumback BA, Wigg M, Ryan L. Synthesis of evidence from epidemiological studies with interval–censored exposure due to grouping. *Biometrics* 2001; **57**(September): 671–80.

187 Shi J, Copas J. Meta-analysis for trend estimation. *Stat Med* 2004; **23**: 3–19.

188 Brumback BA, Holmes LB, Ryan LM. Adverse effects of chorionic villus sampling: a meta-analysis. *Stat Med* 1999; **18**(16): 2163–75.

189 Martin DO, Austin H. An exact method for meta-analysis of case–control and follow-up studies. *Epidemiology* 2000; **11**(3): 255–60.

190 Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press, 1984.

191 Devine EC, Cook TD. Effects of psycho-educational interventions on length of hospital stay: a meta-analytic review of 34 studies. In: Light RJ, ed. *Evaluation Studies Review Annual, vol* **8**. Beverly Hills, CA: Sage, 1983.

192 No authors listed. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995; **345**: 669–85.

193 Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; **323**(July 14): 101–5.

194 Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003; **22**(13): 2113–26.

195 Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979; **86**: 638–41.

196 Iyengar S, Greenhouse JB. Selection models and the file-drawer problem. *Stat Sci* 1988; **3**: 109–17.

197 Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. *Stat Med* 1996; **15**(23): 2493–507.

198 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**(7109): 629–34.

199 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; **50**(4): 1088–101.

200 Dear KBG, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. *Stat Sci* 1992; **7**: 237–45.

201 Hedges LV, Vevea JL. Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. *J Educ Behav Stat* 1996; **21**: 299–332.

202 Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; **60**: 419–35.

203 Givens GH, Smith DD, Tweedie RL. Publication bias in meta-analysis: a Bayesian

data-augmentation approach to account for issues exemplified in the passive smoking debate. *Stat Sci* 1997; **12**: 221–50.

204 Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; **56**(2): 455–63.

205 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358**(3): 252–60.

206 Roest AM, de Jonge P, Williams CD, de Vries YA, Schoevers RA, Turner EH. Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders: a report of 2 meta-analyses. *JAMA Psychiatry* 2015; **72**(5): 500–10.

207 Goldacre B, Turner E, OpenTrials. You can now search FDA approval documents easily at fda.opentrials.net. *BMJ* 2017; **356**: j677.

208 Valsecchi MG, Masera G. A new challenge in clinical research in childhood ALL: the prospective meta-analysis strategy for intergroup collaboration. *Ann Oncol* 1996; **7**(10): 1005–8.

209 Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. *BMJ* 2006; **332**(7553): 1302–8.

210 Baigent C, Keech A, Kearney PM, *et al*. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 2005; **366**(9493): 1267–78.

211 Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. *Med Decis Mak* 2013; **33**(5): 679–91.

212 Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Med Decis Mak* 2013; **33**(5): 671–8.

213 Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Mak* 2013; **33**(5): 657–70.

214 Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Mak* 2013; **33**(5): 641–56.

215 Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity – subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Mak* 2013; **33**(5): 618–40.

216 Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. *Med Decis Mak* 2013; **33**(5): 597–606.

217 Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Mak* 2013; **33**(5): 607–17.

218 Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; **29**(7–8): 932–44.

219 Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med* 2009; **28**(14): 1861–81.

220 Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; **62**(8): 857–64.

221 Signorovitch JE, Wu EQ, Andrew PY, *et al*. Comparative effectiveness without head-t-–head trials. *Pharmacoeconomics* 2010; **28**(10): 935–45.

222 Signorovitch JE, Sikirica V, Erder MH, *et al*. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012; **15**(6): 940–7.

223 Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Mak* 2018; **38**(2): 200–11.

224 Temple RJ. The regulatory evolution of the integrated safety summary. *Drug Inf J* 1991; **25**: 485–92.

225 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; **356**(24): 2457–71.

226 FDA. FDA Drug Safety Podcast: FDA significantly restricts access to the diabetes drug Avandia. www.fda.gov/Drugs/DrugSafety/DrugSafetyPodcasts/ucm227511.htm

227 Bracken MB. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; **357**(9): 937–8.

228 Cepeda MS, Cousins MJ, Carr DB. Musculoskeletal pain. In: Cepeda MS, Cousins MJ, Carr DB, eds. *Fast Facts: Chronic Pain*. Oxford: Health Press, 2007, pp. 70–96.

229 Hauser W, Bernardy K, Uceyler N, Sommer C. Treatment of fibromyalgia syndrome with antidepressants: a meta-analysis. *JAMA* 2009; **301**(2): 198–209.

230 Stone M, Laughren T, Jones ML, *et al*. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ* 2009; **339**: b2880.

231 Webber DM, Montague TH, Bird NP. Meta-analysis of QTc interval-pooling data from heterogeneous trials. *Pharm Stat* 2002; **1**: 17–23.

232 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; **327**: 248–54.

233 Baum ML, Anish DS, Chalmers TC, Sacks HS, Smith H Jr, Fagerstrom RM. A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *N Engl J Med* 1981; **305**(14): 795–9.

234 Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clin Trials* 1997; **18**(6): 568–79; discussion 661–6.

235 Biester K, Lange S. The Multiplicity Problem in Systematic Reviews. https://abstracts.cochrane.org/2005-melbourne/multiplicity-problem-systematic-reviews (accessed May 3, 2019).

236 Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Stat Med* 1997; **16**(24): 2901–13.

237 Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials* 1997; **18**(6): 580–93; discussion 661–6.

238 van der Tweel I, Bollen C. Sequential meta-analysis: an efficient decision-making tool. *Clin Trials* 2010; **7**(2): 136–46.

239 O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**(3): 549–56.

240 Lan KK, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; **70**(3): 659–63.

241 Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. *Res Synth Methods* 2014; **5**(3): 212–20.

242 Imberger G, Wetterslev J, Gluud C. Trial sequential analysis has the potential to improve the reliability of conclusions in

meta-analysis. *Contemp Clin Trials* 2013; **36**(1): 254–5.

243 Miladinovic B, Mhaskar R, Hozo I, Kumar A, Mahony H, Djulbegovic B. Optimal information size in trial sequential analysis of time-to-event outcomes reveals potentially inconclusive results because of the risk of random error. *J Clin Epidemiol* 2013; **66**(6): 654–9.

244 Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med* 2011; **30**(9): 903–21.

245 Thorlund K, Devereaux PJ, Wetterslev J, *et al*. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *Int J Epidemiol* 2009; **38**(1): 276–86.

246 Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008; **61**(8): 763–9.

247 Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008; **61**(1): 64–75.

248 Lan KKG, Hu M, Cappelleri JC. Applying the law of iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Stat Sin* 2003; **13**(4): 1135–45.

249 Hu M, Cappelleri JC, Lan KK. Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clin Trials* 2007; **4**: 329–40.

250 Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol* 1995; **48**(1): 45–57; discussion 59–60.

251 Ziegler RG, Mayne ST, Swanson CA. Nutrition and lung cancer. *Cancer Causes Control* 1996; **7**(1): 157–77.

252 Hennekens CH, Buring JE, Manson JE, *et al*. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 1996; **334**(18): 1145–9.

253 Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004; **364**(9450): 2021–9.

254 Gandey A. Medscape Medical News: Merck fights back. www.medscape.com/viewarticle/538074 (accessed May 3, 2019).

255 Ryan C, Leonardi CL, Krueger JG, *et al*. Association between biologic therapies for chronic plaque psoriasis and cardiovascular events: a meta-analysis of randomized controlled trials. *JAMA* 2011; **306**(8): 864–71.

256 Tzellos T, Kyrgidis A, Zouboulis C. Re-evaluation of the risk for major adverse cardiovascular events in patients treated with anti-IL-12/23 biological agents for chronic plaque psoriasis: a meta-analysis of randomized controlled trials. *J Eur Acad Dermatol Venereol* 2013; **27**(5): 622–7.

257 Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000; **10**: 325–37.

258 Turner E, Moreno SG, Sutton AJ. Ranking antidepressants. *Lancet* 2009; **373**(9677): 1760–62.

259 Yahav D, Paul M, Fraser A, Sarid N, Leibovici L. Efficacy and safety of cefepime: a systematic review and meta-analysis. *Lancet Infect Dis* 2007; **7**(5): 338–48.

260 Solomon SD, McMurray JJ, Pfeffer MA, *et al*. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* 2005; **352**(11): 1071–80.

261 Proschan MA, Lan KKG, Wittes JT. Monitoring for safety. In: Proschan MA, Lan KKG, Wittes JT, eds. *Statistical Monitoring of Clinical Trials*. New York: Springer, 2010, pp. 155–74.

262 Song F, Altman DG, Glenny AM, Deeks J. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003; **326**(March 1): 1–5.

263 Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network – improving the evidence of medical-product safety. *N Engl J Med* 2009; **361**: 645–7.

264 US Department of Health and Human Services. The Sentinel Initiative: A National Strategy for Monitoring Medical Product Safety. www.fda.gov/files/about%20fda/published/The-Sentinel-Initiative-May-2008-Report.pdf (accessed May 3, 2019).

265 US Department of Health and Human Services. FDA's Sentinel Initiative – Ongoing Projects. www.fda.gov/safety/fdas-sentinel-initiative/fdas-sentinel-initiative-ongoing-projects (accessed May 3, 2019).

37

Validity of Drug and Diagnosis Data in Pharmacoepidemiology

Mary Elizabeth Ritchey[1], Suzanne L. West[2,3], and George Maldonado[4]

[1] *RTI Health Solutions, RTI International, Research Triangle Park, NC, USA*
[2] *RTI International, Research Triangle Park, NC, USA*
[3] *Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA*
[4] *Division of Environmental Health Sciences, School of Public Health, University of Minnesota, Minneapolis, MN, USA*

To provide evidence-based care, clinicians need to know the benefits and risks of the medications they are prescribing (see Chapter 35), and this information needs to come from robust research. For example, evidence for medication efficacy typically comes from randomized controlled trials, whereas establishing the magnitude of a drug safety concern often comes from observational studies using self-reported data or from electronic data such as administrative claims data or electronic health records (EHRs). Previous editions of this chapter focused primarily on self-reported data or administrative claims, but with the growing availability of EHR data, pharmacoepidemiologists are increasingly using these data for research because they contain more granular information such as the reason for medication prescription, laboratory test results, and patient vitals (e.g., blood pressure and weight).

Clinical Problems to be Addressed by Pharmacoepidemiologic Research

Of particular concern to the subject of this book is the validity of data on drug exposure and disease occurrence because the typical focus of pharmacoepidemiologic research is often the association between a medication and an adverse drug event. Further, many potential confounders of importance in pharmacoepidemiologic research (although certainly not all) are either drugs or diseases. Clinicians recognize that patients very often do not know the names of the drugs they are taking currently. Thus, it is a given that patients have difficulty recalling past drug use accurately, at least in the absence of aids to enhance recall. Superficially at least, patients cannot be considered reliable sources of diagnosis information either; in some instances, they may not have even been told

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

the correct diagnosis, let alone recall it. Yet, these data elements are crucial to pharmacoepidemiologic studies that ascertain data using questionnaires. Special approaches have been developed by pharmacoepidemiologists to obtain such data more accurately when using self-report for data collection, but the success of these approaches needs to be considered in detail.

Besides self-reported data, pharmacoepidemiologists have been using administrative claims data for more than 30 years to evaluate drug safety. We discuss validity issues with using these data for research. However, the changing landscape of healthcare requires reassessing the validity of the data pharmacoepidemiologists are now using for their research and how these data impact clinical practice.

More and more, pharmacoepidemiologists are turning to EHR data for their research.

Whereas the increased granularity of EHR data is a benefit for their use in pharmacoepidemiology, important limitations of these data include their potential incompleteness and lack of interoperability across health systems. Unless EHR data arise from "closed" healthcare systems where patients receive all their outpatient and inpatient care, then the EHR data may represent only a portion of the patients' health problems and care received. If EHR data from multiple health systems are used, even if the health systems use the same EHR vendor, the data may need to be restructured so that they are consistent across all data arising from all health systems. The clinician reviewing evidence for patient care that arises from studies using EHR data trusts that these data have been curated sufficiently to produce robust and valid study findings.

Methodologic Problems to be Solved by Pharmacoepidemiologic Research

There are five major methodologic problems associated with validity of data for pharmacoepidemiologic research: indices of measurement error, quantitative measurement of validity, quantitative measurement of reliability, measurement error in pharmacoepidemiologic research, and adjusting measures of association for measurement error.

Indices of Measurement Error

Two main comparisons may be drawn between two (or more) methods of data collection or sources of information on exposure or outcome: validity and reliability. Many different terms have been used to describe each, resulting in some confusion. Although the literature uses the term *validation study* or *verification* to describe the agreement between two sources of information, *concordance* or *agreement* might be a more appropriate term to describe the comparison between data sources because validation requires a "gold standard." In the following discussion, we define and differentiate between validity and reliability. Validity is assessed using sensitivity and specificity, while reliability is typically measured using percent agreement and kappa [1].

Quantitative Measurement of Validity

Only when one of the methods or sources is clearly superior to the other can the comparison be said to measure validity. The superior method or source is often called a "gold standard." In recognition that a method or source can be superior to another method or source without being perfect, the term *alloyed (or tarnished) gold standard* has been used [2].

For a binary exposure or outcome measure, such as "ever" versus "never" use of a particular drug, two measures are used to assess validity. Sensitivity measures the degree to which the potentially inferior source or method correctly identifies individuals who, according to the superior method or source, possess the characteristic of interest (i.e., ever used the drug). Specificity measures the degree to which the inferior source

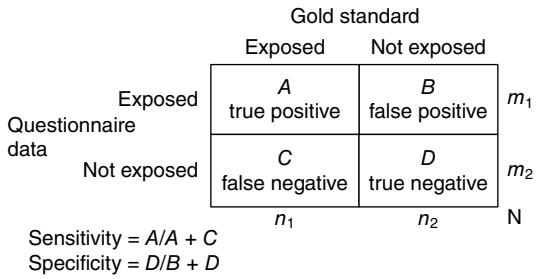


Figure 37.1 Formulas for calculating sensitivity and specificity.

or method correctly identifies individuals who, according to the superior method or source, lack the characteristic of interest (i.e., never used the drug). Figure 37.1 illustrates the calculation of sensitivity and specificity.

Sensitivity and specificity are the two sides of the validity coin for a dichotomous exposure or outcome variable. In general, sources or methods with higher sensitivity tend to have lower specificity, and methods with higher specificity tend to have lower sensitivity. In these very common situations, neither of the two sources or methods compared can be said to have superior overall validity. Depending on particulars of the study setting in which the research question is addressed, either sensitivity or specificity may be the more important validity measure. Moreover, absolute values of these measures can be deceiving. For instance, if the true prevalence of ever use of a drug is 5%, then an exposure classification method or information source with 95% specificity (and perfect sensitivity) will incorrectly double the measured prevalence to 10%. The ultimate criterion of importance of a given combination of sensitivity and specificity is the degree of bias exerted on a measure of effect such as an estimated relative risk due to measurement error.

As measures of validity, sensitivity and specificity have "truth" (i.e., the classification according to a gold standard or an alloyed gold standard) in their denominators. Investigators should take care not to confuse these measures with the predictive values of positive and negative classifications, which include the inferior measure in their denominators. We distinguish here between the persons who *actually* do or do not have an exposure or outcome with those who are *classified* as having it or not having it (using the potentially inferior or alternative data source). The proportion of persons classified as having the exposure or outcome who truly do have the exposure or outcome is the positive predictive value. The proportion of persons correctly classified as lacking the exposure or outcome is the negative predictive value.

Assessment of the positive predictive value (as is performed in many validation studies in administrative claims and EHR data) of an outcome does not directly measure the validity of the data source. Predictive values are measures of performance of a classification method or information source, not measures of validity. Predictive values depend not only on the sensitivity and specificity (i.e., on validity) but also on the prevalence of the exposure or outcome. Thus, if a method or information source for classifying persons with respect to outcome or exposure has the same sensitivity and specificity in two populations but those populations differ in their outcome or exposure prevalence, the source or method will have different predictive values in the two populations. Nonetheless, all measures are useful and the most important one will depend on the question being answered. Ideally, one would design a validation study to calculate sensitivity and specificity as well as positive and negative predictive values.

In some validation studies, one method or source may be used as a gold standard or as an alloyed gold standard to assess another method or source with respect to only one side of the validity coin. Studies that focus on the completeness of one source, such as studies in which interview responses are compared with prescription dispensing records to identify drug exposures that were forgotten or otherwise not reported by the respondents, may measure (more or less accurately) the sensitivity of the

interview data. However, such studies are silent on the specificity unless one acknowledges strong assumptions (e.g., that the respondent could not have obtained the drug in a way that would not be recorded in the prescription dispensing records). Similarly, in administrative claims data, prescriptions that are filled outside the insurance plan may not be captured in the database, especially for generic drugs that are less costly to purchase outright rather than using a co-pay.

For a drug exposure, a true gold standard would be a list of all drugs the study participant has taken (i.e., ingested), including dose, duration, and dates of exposure. This drug list might be a diary of prescriptions the study participants kept or, perhaps more readily available, a computerized database of filled prescriptions, although neither of these data sources is a genuine gold standard. Prescription diaries cannot be assumed to be kept in perfect accuracy. For instance, participants may tend to record that drug use was more regular and complete than it actually was or that use adhered to the prescribed regimen. Similarly, substantial gaps may exist between the point at which a prescription is filled and when it is ingested, if it is ingested at all. See Chapter 38 for further discussion of adherence.

Two methods are used to quantify the validity of continuously distributed variables, such as duration of drug usage. The mean and standard error of the differences between the data in question and the valid reference measurement are typically used when the measurement error is constant across the range of true values (i.e., when measurement error is independent of where an individual's true exposure falls on the exposure distribution in the study population) [3]. With the caveat that it is generalizable only to populations with similar exposure distributions, the product–moment correlation coefficient may also be used.

High correlation between two measures does not necessarily mean high agreement. For instance, the correlation coefficient could be very high (i.e., close to 1), even though one of the variables systematically overestimates or underestimates values of the other variable. The high correlation means that the over- or underestimation is systematic and very consistent. When the two measures being compared are plotted against each other and they have the same scale, full agreement occurs only when the points fall on the line of equality, which is 45° from either axis [4]. However, perfect correlation occurs when the points lie along any straight line parallel to the line of equality. It is difficult to tell from the value of a correlation coefficient how much bias will be produced by using an inaccurate measure of disease or exposure.

Quantitative Measurement of Reliability

When the same data collection method or source of information is used more than once for the same information on the same individual, comparisons of the results measure the reliability of the method or information source. An example of a reliability study is a comparison of responses in repeat interviews using the same interview instrument. Reliability is not validity, although the term is sometimes used, inaccurately, as such. In general, studies that measure mere agreement are all too commonly interpreted as though they measured validity or accuracy. The term *reliability* tends to be used far too broadly to refer variously not only to reliability itself, but to agreement or validity as well. Researchers and others should take greater care with the way they use such terms.

When different data collection methods or different sources of information are compared (e.g., comparison of prescription dispensing records with interview responses), and neither of them can be considered distinctly superior to the other, the comparisons measure mere agreement. Agreement between two sources or methods does not imply that either is valid.

To evaluate reliability or agreement for categorical variables, the percentage agreement between two or more sources and related (kappa) coefficient are used. They are used only when two imperfect classification schemes are being compared, not when one classification method may be considered *a priori* superior to the other [3,5]. The kappa statistic is the percentage agreement corrected for chance [3]. Agreement is conventionally considered poor for a kappa statistic less than zero, slight for a kappa between zero and 0.20, fair for a kappa of 0.21–0.40, moderate for a kappa of 0.41–0.60, substantial for a kappa of 0.61–0.80, and almost perfect for a kappa of 0.81–1.00 [1]. Figure 37.2 illustrates the percentage agreement and kappa calculations for a reliability assessment between questionnaire data and medical record information.

The intraclass correlation coefficient is used to evaluate the reliability of continuous variables [5]. It reflects both the average differences in mean values as well as the correlation between measurements. The intraclass correlation coefficient indicates the degree to which the total measurement variation is due to the differences between the subjects being evaluated and to differences in measurement for one individual. When the data from two sets of measurements are identical, the intraclass correlation coefficient equals 1.0. Under certain conditions, the intraclass correlation coefficient is exactly equivalent to Cohen's weighted kappa [3]. It is impossible to translate values of measures of agreement, such as kappa, into expected degrees of bias in exposure or disease associations.

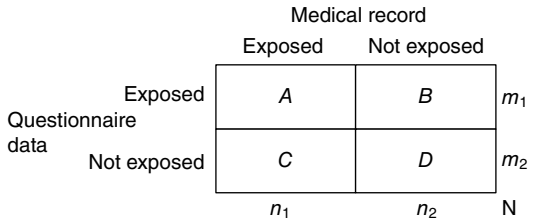


Figure 37.2 Formulas for calculating the percent agreement and K.

Measurement Error in Pharmacoepidemiologic Research

Epidemiologic assessments of the effects of a drug on disease incidence depend on an accurate assessment of the study exposure, disease occurrence, and variables to be adjusted in the statistical analysis. Measurement error for any of these factors may incorrectly identify a risk factor in the study that does not exist in the population or, conversely, may fail to detect a risk factor when one truly exists.

In an epidemiologic study, the measure of association is often based on the number of subjects categorized by the cross-classification of presence or absence of disease and exposure. For example, when questionnaire data are used to study the association between drug A and disease B, study participants who forget their past exposure to drug A would be incorrectly classified as nonexposed. Similarly, if a provider uses a diagnosis code to document the process of testing and ruling out a disease and then a researcher uses the diagnosis code as a study outcome, then the person would be incorrectly classified as having the outcome. This misclassification is a measurement error. Although the measurement process often involves some error, if this measurement error is of sufficient magnitude, the validity of the study's findings is diminished.

Surprisingly, measurement error is often ignored in epidemiologic studies. Jurek *et al.* [6] reported the results of a random survey of studies published in three major epidemiology journals; they concluded the following for exposure-measurement error (EME): "Overall, the potential impact of EME on error in epidemiologic study results appears to be ignored frequently in practice" (p. 871).

Measurement error is a potentially serious cause for concern in epidemiologic studies, and therefore, for several reasons, this should not be ignored when analyzing and interpreting pharmacoepidemiologic study results. First, small amounts of measurement error can cause large amounts of error in study results. For example, consider a pharmacoepidemiologic study of nonsteroidal antiinflammatory drug (NSAID) A versus NSAID B on gastrointestinal (GI) bleed (Figure 37.3). In a study with a total number of study subjects equal to more than 22 000, if only 10 subjects are misclassified with respect to their exposure or disease (five who actually took NSAID B are incorrectly classified as having taken NSAID A, and five users of NSAID A without GI bleed are incorrectly classified as having GI bleed), the observed odds ratio (OR) would be 2.1 when the correct OR is in fact 1.0.

Second, measurement error can cause study results to overestimate or underestimate true effect sizes, and there is no simple rule for predicting the direction of the error in real-life situations. We now understand that these old and often-cited heuristics are not necessarily true, except under special conditions that are not likely to occur in practice: (1) nondifferential misclassification always produces bias toward the null, and (2) bias toward the null always produces an observed relative risk that is an underestimate of

Case-Control Study of NSAIDs and Myocardial Infarction (MI)

No exposure misclassification

MI cases recall exposure just as well as those without MI

	NSAID use	No NSAIDs	
MI	200	200	OR = 2.5
No MI	240	600	

MI cases recall exposure better than those without an MI

MI	Sensitivity=	0.95
	Specificity=	0.9
No MI	Sensitivity=	0.8
	Specificity=	0.7

	NSAID use	No NSAIDs	
MI	210	190	OR = 1.4
No MI	372	468	

MI cases do not recall exposure as well as those who did not have an MI

MI	Sensitivity=	0.8
	Specificity=	0.9
No MI	Sensitivity=	0.9
	Specificity=	0.7

	NSAID use	No NSAIDs	
MI	180	220	OR = 0.9
No MI	396	444	

Figure 37.3 Example of differential misclassification of exposure.

the true relative risk. These heuristics are unlikely to be true in practice for the following reasons.

- Conditions beyond nondifferentiality are required to guarantee bias is toward the null [7–12] (e.g., when the degree of exposure measurement error systematically differs across levels of a polychotomous or continuous exposure variable, or when errors in measuring the exposure and outcome are not independent).
- Even when the above conditions beyond nondifferentiality are met, exact nondifferentiality is required to guarantee bias is toward the null [13,14].
- Also required to guarantee bias is toward the null is either (1) the absence of other study biases (e.g., absence of confounding, absence of bias due to nonrandom subject selection/participation) or (2) the combined effect of all other biases is also toward the null [13].
- Bias is a statistical term that is defined as the difference between the true value and the expected value of an estimator (i.e., the average of study results over hypothetical repetitions of the study). Bias is not the difference between the observed estimate for one repetition of the study and the true value. This important distinction was not appreciated in earlier writings on this topic, and even today we epidemiologists are not careful in our use of the term bias. Therefore, when bias is toward the null, the expected value of the estimator is shifted toward the null, but an observed estimate can be an overestimate of the true relative risk due to the influence of random error [13]. (Similarly, when there is no bias of any kind, one observed estimate can be an overestimate or an underestimate of the true relative risk simply due to random error.)

Third, error in measuring variables to be adjusted in the analysis can result in only partial adjustment for the mismeasured variables [15].

Adjusting Measures of Association for Measurement Error

One can use sensitivity analysis methods (also known as uncertainty analysis and bias analysis) [16–24] to adjust measures of association for measurement error as well as for other study biases. (As used in this context, the meaning of the term *sensitivity* differs from its other epidemiologic meaning as the counterpart to specificity as a measure of classification validity.) Sensitivity analysis is the last line of defense against biases after every effort has been made to eliminate, reduce, or control them in study design, data collection, and data analysis. In a sensitivity analysis, one alters key assumptions or methods reasonably to see how sensitive the results of a study are to those variations. (See Chapter 38 for discussion of sensitivity analyses in pharmacoeconomic studies.)

One key assumption, usually implicit, in any study that does not quantitatively account for the possibility of error in measuring the study exposure or study outcome is that the exposure and the outcome in a study have been measured accurately. With estimates of sensitivity and specificity from validation studies (from previous research or from a subsample within the study analyzed) or "guesstimates" from expert experience and judgment, one can modify this assumption and use sensitivity analysis methods to "back calculate" what the results might have looked like if more accurate methods had been used to classify participants with respect to outcome, exposure, or both [17,25].

For many years, a qualitative and informal version of this kind of assessment has been conducted. However, the net result is controversy, with investigators judging the bias small and critics judging it large. Further, in the absence of a formal bias analysis, intuitive judgments, even those of the most highly trained and widely experienced investigators, can be poorly calibrated in such matters. Formal sensitivity analysis makes the assessment of residual bias transparent and

quantitative and forces the investigator (and other critics) to defend criticisms that in earlier times would have remained qualitative and unsubstantiated. An important and well-known historical example is the bias from nondifferential misclassification of disease proposed by Horwitz and Feinstein [26] to explain associations between early exogenous estrogen preparations and endometrial cancer. When proper sensitivity analyses were conducted to assess this bias, only a negligible proportion of those associations were explained by bias [26–28].

Epidemiologic applications of quantitative methods with a long history in the decision sciences have become accessible for quantifying uncertainties about multiple sources of systematic error in a probabilistic manner [24,29–31]. These methods permit the incorporation of available validation data as well as expert judgment about measurement error, uncontrolled confounding, and selection bias along with conventional sampling error, and prior probability distributions for effect measures themselves, to form uncertainty distributions. These approaches have been used practically in pharmacoepidemiology studies such as in assessing selection bias in a study of topical coal tar therapy and skin cancer among severe psoriasis patients [30]; exposure misclassification and selection bias in a study of phenylpropanolamine use and stroke [24]; and selection bias, confounder misclassification, and unmeasured confounding in a study of less than standard therapy and breast cancer mortality [29], as well as in other clinical and nonclinical applications [18,31–39].

Sometimes biases can be shown to be of more concern and sometimes of less concern than intuition or simple sensitivity analysis might suggest. Almost always, the probabilistic uncertainty about these sources of systematic error dwarfs the uncertainty reflected by conventional confidence intervals (CIs). By the use of these methods, the assessment of systematic error can move from a qualitative discussion of "study limitations," beyond sensitivity analyses of one scenario at a time for one source of error at a time, to a comprehensive analysis of all sources of error simultaneously. The resulting uncertainty distributions can not only supplement but also supplant conventional likelihood and *P* value functions, which reflect only random sampling error. As a result, much more realistic, probabilistic assessments of total uncertainty attending to effect measure estimates are in the offing [19].

Currently Available Solutions

Conducting Validation Studies to Assess Self-Reported Data

In 1979, Leon Gordis commented that epidemiologists have become so enamored with analyzing their data that they have paid too little attention to the validity of the raw data being analyzed with these sophisticated techniques [40]. Gordis' comment reflects a time when pharmacoepidemiologic research was typically conducted by using questionnaires to gather data. The field was only just beginning to use data that arose from the provision of healthcare, including health insurer data such as Medicaid claims.

This section of the chapter focuses on the collection and validation of self-reported data for pharmacoepidemiologic research. We begin this section with a brief discussion of how individuals store and retrieve information from memory, tasks that are required when responding to a questionnaire. We use an example of how a person might recall a depression episode to illustrate retrieval of specific information from memory. Recognizing the challenges of information retrieval, we discuss best practices for designing questions to elicit specific drug and diagnosis information. Separately for drugs and diagnoses, we discuss the influence of comparator selection when validating self-reported data, the accuracy of recall, and the factors

influencing recall and provide examples for illustration.

Autobiographical Memory and the Response Process

Pharmacoepidemiologic research that relies on self-reported data requires asking study respondents to recall events or exposures that occurred at some time in the past, with recall intervals spanning from days to years. The types of temporal questions study respondents are often asked and that require the memory processes are as follows [41]:

- Time of occurrence, which requires respondents to provide a date when an event occurred, such as when they were diagnosed with a particular condition.
- Duration questions such as, "How long did you take drug A?"
- Elapsed time, which asks how long it has been since an event occurred, including questions such as, "How many months has it been since you last took drug A?"
- Temporal frequency questions that ask respondents to report the number of events that occurred over a specific time period, such as "How many visits did you make to your primary care practitioner in the past six months?"

To appreciate the accuracy of data derived by respondent recall for addressing these types of questions, it is important to understand how we process, organize, and recall autobiographical information, which is key to the response process. Creating and retrieving information from autobiographical memories is a three-step process. Information that comes in via sensory or emotional input (e.g., visual, hearing, semantic) is *encoded* into a construct that can be stored within the brain. The next step is *storage*, which refers to how the brain retains the information, typically in either short- or long-term memory. *Retrieval or recall* of memories requires reaccessing information that was previously encoded and stored. Recall effectively returns a memory from long-term storage to short-term memory, where it can be accessed for retrieval purposes [42,43]. Current thinking is that retrieval of information from autobiographical memory is goal oriented, where the retrieval process requires bringing together spatial, temporal, and social information with information derived from the emotions and senses [42].

The recall of encoded or catalogued information from memory is thought to be facilitated by using important personal milestones [41]. Thus, when respondents are asked to recall a visit to a doctor that may have occurred at a particular point in time, researchers believe that the respondents use scripts (a generic mental representation of the event) to help retrieval. For example, the respondent first contemplates a doctor visit in general and then supplements this script with details relevant to the particular visit that require contemplation for specific criteria (e.g., diagnosis) and timing (e.g., a particular year). In general, underreporting of medical conditions and health visits is more widespread as the interval since the event increases [44–46].

Recent evidence suggests that age affects memory details, with older individuals recalling slightly more details than younger individuals. Using an instrument focused on words used in everyday spoken and written language to measure autobiographical memory, Gardner and colleagues noted that recall of content and details for events and objects was slightly greater in adults 46–78 years old compared to those 26–45 years of age for both recent and remote memories [47]. There was little difference between the two age groups for recalling individuals and temporal details of events.

Applying what we know about how autobiographical memory is organized and the recall process in general helps us to understand survey response. A respondent undergoes four key tasks when asked to answer a questionnaire: (1) question comprehension and interpretation, (2)

search for and retrieval of information to construct an answer to the question depending on whether appropriate cues are given, (3) judgment to discern the completeness and relevance of memory for formulating a response, and (4) development of the response based on retrieved memories [41,48–50]. If survey instrument developers pay too little attention to the first two key tasks, their questions can be too vague or complex for respondents to marshal retrieval processes appropriately.

The following example best illustrates the response process [41] for recalling the date on which a respondent's depression was diagnosed (January 2015). The recall process begins with the respondent being uncertain whether the depression was diagnosed in 2014 or 2015. To work towards identifying the correct year, the respondent recalls that the depression occurred after he lost his job. The job loss was particularly traumatic because he and his wife just purchased their first home a few months previously, and now, with the loss of his income, they were at risk of losing the house. The home purchase was a landmark event for this respondent, and he remembers that it occurred in mid-2014, just as their children finished the school year. So, in 2014 he lost his job, near the end of the year because the holiday season was particularly grim. He remembers that his depression was diagnosed after the holidays, but was it January or February of 2015? It was January 2015 because he was already taking antidepressants by Valentine's Day, when he went out to dinner with his wife and he could not drink wine with his meal. This chronology is illustrated in Figure 37.4. We describe below how to use the response process to design questions to elicit the self-reported information requested.

As illustrated in Figure 37.4, landmark events probably serve as the primary organizational units of autobiographical memory and, as such, anchor information retrieval [51]. In particular, the example shows how the respondent used landmark and other notable events, relationships among datable events, and general knowledge (holiday period and children finishing the school year) to reconstruct when his major depression

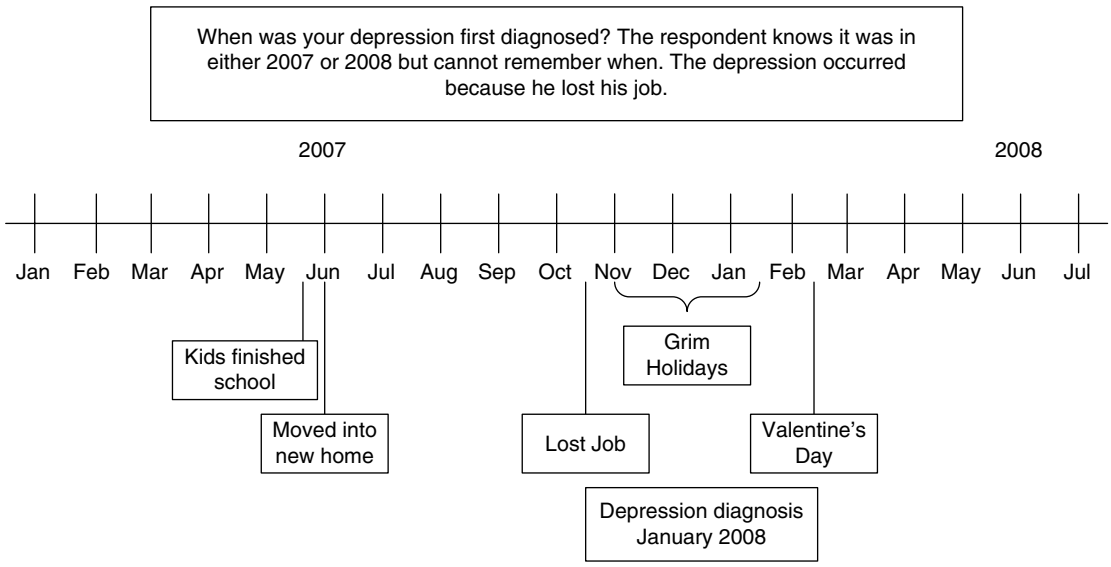


Figure 37.4 Recall schematic for showing how date of depression diagnosis was determined.

was first diagnosed. An important caveat is that this respondent was willing to expend considerable effort to search his memory to determine when his depression was diagnosed – this may not be the situation for all respondents.

The next section takes what we know about autobiographical memory and the response process to develop questionnaires for pharmacoepidemiologic research.

Best Practices for Questionnaire Design

Designing a questionnaire for collecting pharmacoepidemiologic data requires consideration of the challenges and limitations of autobiographical memory as described above and careful planning and pretesting [52] before fielding the study. Survey researchers encourage use of several general techniques to assist respondents in recalling information accurately, including use of reference periods (e.g., "in the past 12 months, that is, since December 1, 2017, how many times did you…"), event histories and calendars like the one in Figure 37.4, diaries, and photos of medications. We provide a more in-depth discussion of questionnaire design for collecting medication and diagnosis data later. We recommend that, after collecting the self-reported data using techniques to maximize their accuracy, and prior to the analysis, researchers assess their accuracy for addressing the study hypothesis by comparing the data to another data source such as health insurer claims or EHRs [53].

We suggest the following steps be considered during the design and initial analysis stages of a study requiring data collection via questionnaire.

- Use validated instruments or validated questions whenever possible.
- Consider question banks if new questions are required, such as World Bank's Living Standards Measurement Study [54] and Q-Bank [55,56].
- Use question assessment tools to determine the likelihood of response error. These tools include the Question Appraisal System [57], the Survey Quality Predictor (SQP) [58], and the Question Understanding Aid (QUAID) [59,60].
- Strive for a fifth-grade literacy level if you must develop new survey questions to be used for a general population [61].
- Pretest the questions using cognitive testing [62–64] to assess respondent comprehension of new questions.

The process of *satisficing* occurs when respondents expend the least psychological and emotional effort possible to provide an acceptable answer to a survey question rather than an optimal answer [65,66]. To minimize satisficing, questionnaire developers should consider the length of the instrument and the number of response categories. When faced with a long list of choices and depending on the mode of questionnaire administration (i.e., telephone versus self-administered), respondents may choose answers from either the top or the bottom of the list to minimize effort. For this reason, it is often recommended to randomize response options. Respondents with lower cognitive skills and less education, when challenged with discerning the best possible response, are more apt to settle for a satisfactory rather than an optimal response. Because accuracy of response is critical for pharmacoepidemiologic research, questionnaire developers must consider methods to minimize response burden leading to satisficing.

With the increasing availability of broadband and the population's access to the internet, more surveys are moving away from face-to-face and telephone interviewer administration to web-based surveys. This modality requires the same considerations for question design as described earlier, but because no interviewer is available, usability should be tested as well. Usability evaluates the survey–respondent interaction: essentially, how efficiently and effectively respondents can answer the web-based questions [67,68]. For example, usability evaluates screen size, button placement, and formatting issues specific to web

applications, especially for questionnaires using mobile technologies. Usability assessments can be combined with other pretesting modalities [52], including cognitive interviews, by embedding probes that allow respondents to explain why they provided their answers [69], a parallel to face-to-face cognitive testing without requiring an interviewer [52].

The earlier discussion focused on measurement error related to survey design and to respondent motivation. Measurement error can also be attributed to improper training of interviewers and poor data entry quality. The degree to which one understands the measurement error associated with key variables critical to the analysis can be assessed by using several different modelling approaches, which Biemer discusses in more detail [70].

Assessing the Accuracy of Self-Reported Data

Despite researchers using the best methods for designing questionnaires to elicit specific information on medications used previously and past diagnoses, self-reported data still require evaluation for accuracy to ensure valid findings. Ideally, researchers will have access to a truly accurate comparison source (i.e., gold standard) so that sensitivity and specificity can be calculated for use in bias analyses. For example, we can use pill counts, chemical markers inserted into the pills, electronic monitoring caps, or pharmacy dispensing databases to assess self-reported medication use. As discussed earlier, depending on the comparison data source, it may only be possible to calculate either sensitivity or specificity

Methodologic studies that use alternative data sources, such as prospectively collected drug data (e.g., from diaries), or databases of dispensed drugs can measure both sensitivity and specificity if one assumes that these databases are true gold standards. In pharmacoepidemiology, lower sensitivity is often more of a concern than lower specificity. Questionnaires that underreport diseases or miss drug exposures because the medication was filled without using the pharmacy plan (e.g., when the co-pay is higher than the cost of the medication) – that is, data sources with low sensitivity – cannot be used to rigorously evaluate drug–disease associations. Alternatively, low specificity is often less of a problem in pharmacoepidemiology unless the characteristic with low specificity also has very low prevalence in the population being studied. For example, because the incidence of Stevens–Johnson syndrome is low, a small degree of misclassification when using administrative claims data in which the case definition uses the *International Classification of Diseases, Ninth Edition, Clinical Modification* (ICD-9-CM) code 695.1 will include several skin problems other than Stevens–Johnson (i.e., the false-positive rate would be high) [71].

Besides the need for completeness on the individual level, the comparator database must have information for all persons whose information is to be assessed for accuracy. Systematic omissions of specific population groups, such as certain ethnic or racial groups, diminish the quality of the database.

In the next section of the chapter, we discuss issues in using the medical record as a comparator data source to evaluate the accuracy and completeness of survey data on medication and diagnoses ascertained via self-report. We discuss use of automated databases as a comparator data source for assessing validity and reliability of self-reported information in a later section.

Influence of Comparator Selection for Assessing Self-Report Accuracy

The early work on evaluating the completeness of self-reported diagnosis and medication data typically used paper medical records for comparison [72–74]. In summary, several studies from the late 1980s through 2000 indicated that inpatient medical records were often missing outpatient medications [75–77]. Similarly, outpatient

medical records were also often incomplete, and completeness varied by the number and type of medication the patients were taking [78–82]. Diagnoses or other relevant inpatient information were often omitted from patient records as well [83–86]. These studies indicate that the paper medical record may not be that useful for validation of diagnosis and medication data. With the greater availability of EHR software and policy levers incentivizing their use, providers and hospitals in the United States have been moving to EHRs, making paper medical records obsolete.

Nonetheless, regardless of whether the medical record is paper or electronic, one needs to understand its availability, completeness, and accuracy to determine whether it is adequate for evaluating the accuracy of self-reported information. Retrieval of medical records depends not only on a person's ability to remember and report who prescribed the drug or diagnosed the condition in question, but on whether the healthcare provider recorded the information (and recorded it accurately) and on the availability of the medical record for review. If the medical record cannot be retrieved because the healthcare provider could not be identified, the provider had retired, or the record was destroyed or lost, the events cannot be verified.

While paper medical records are often incomplete, how complete are EHR data for assessing the accuracy of self-reported diagnosis and medication data? This question requires reframing to consider EHR completeness at both the individual patient and the institution level. In the US, healthcare is fragmented. Patients see multiple providers, are treated in several different health settings (e.g., chiropractors, podiatrists), and may become inpatients at several different hospitals [87,88]. Thus, accessing patients' outpatient and inpatient medical records does not guarantee that a researcher will have all medical care provided and drugs prescribed to the patient. For example, if a researcher is able to access only the patient's primary care records, it is possible that the results of cardiology tests to confirm a diagnosis or medications for that diagnosis will not be available. However, when patients are seen by integrated delivery systems that include primary care, multiple specialties, and inpatient care, there is a greater likelihood that the EHR will contain most of the care provided and medications prescribed to the patient.

In addition, the EHR data themselves may be no more accurate than paper records if the EHR data simply substitute for paper records. For example, exposure information about medications or important confounders (e.g., smoking) may be incomplete if clinicians do not ascertain this information and correctly enter it into the EHR. Another problem introduced by EHRs is the potential for errors inherent to electronic data entry, such as copying and pasting of incorrect data from other parts of the record, of expired or irrelevant clinical information, or of incorrect and/or unverified medication lists [89].

Self-Reported Drug Data From** De Novo **Questionnaire Studies

This section summarizes what is known on how well respondents recall prescription and over-the-counter medication use, factors that influence recall, such as the type of medications being queried, as well as questionnaire design features suggested to improve recall accuracy.

Accuracy and Recall

Several studies have evaluated self-reported recall accuracy for current or past medication use compared with prospectively collected cohort data or pharmacy, hospital, and outpatient medical record documentation. Overall, published studies indicate that people accurately remember ever using a medication and when they first began using some medications, although they do not remember brand names and duration of use as well [90–98]. Current use of chronically used medications, such as statins, beta-blockers, and calcium channel blockers,

was recalled with ≥95% sensitivity and specificity when a mailed medication inventory was compared to pharmacy records [99]. In general, greater inaccuracies have been noted as more time elapsed between occurrence of exposure and its subsequent reporting [91,95,97]; this tendency was especially true for over-the-counter NSAID use in contrast with prescription NSAID use for recall over a two-month period [100].

Accuracy of self-reporting medication use varies by several factors. For example, chronically used medications (especially those with more refills) are recalled more often than acute exposures, as are the first and most recent brands in a class; a person recalls multiple medications in one class more frequently than single medication exposure; and salient exposures (those that prompted study initiation) are more accurately recalled than common and less disconcerting exposures [90,91,96,101–105]. For prescription drugs, recall between self-reported use and medical records was moderately accurate, but for over-the-counter medications and vitamin supplements, accurate recall was poor [106]. Discrepancies are due to both underreporting (e.g., respondent forgot medication was taken) and underdocumenting (e.g., physician was unaware of medication use or did not record patient's use in chart) [79–81 92,101,103,105–108] and differed by therapeutic class [106,107,109–116]. When self-reported data were compared to multiple sources (e.g., medical records and pharmacy dispensing), verification for self-reported use was higher than that for a single source [117].

Influences on Accuracy

Influence of Questionnaire Design

As reported in a systematic review, several factors affect the accuracy of medication exposure reported via questionnaire [118]. Researchers can facilitate recall and reporting of medication use by indication-specific questions, memory prompts (such as drug photo), a list of drug names, or a calendar to record life events [70–74,93,119]. Medication-specific or indication-specific questions can identify most medications respondents are currently using, rather than a general medication question such as, "Have you taken any other medications?" [105]. Similarly, open-ended questions such as, "Have you ever used any medications?" yielded less than half of the affirmative responses for use of three different medications [120]. Using the filter question "Did you use any medications in the three months before or during your pregnancy?", van Gelder and colleagues noted that many women failed to report medications that they had been dispensed for pain or infections. These findings could be attributed to poor recall, but they may also be due to women having chosen not to take the dispensed medications [121]. If researchers choose to use open-ended medication questions, adding indication-specific questions that facilitate recall of medication exposures may be useful. Finally, 20–35% of respondents reported drug exposure only when asked medication (name)-specific questions [120].

Response order may affect recall, as noted with malaria medications when respondents had more than one episode of malaria [122]. Regardless of how frequently the medication is used for treating malaria in general, medications listed earlier in the response set tended to be selected more frequently than those listed later – a finding that may be related to satisficing, as discussed earlier [65].

A comparison of self-report of current and recent medication use (within the past two years) to pharmacy records of dispensed prescriptions for multiple drug classes found that the number of drug dispensings recalled was highest for cardiovascular medications (66%) and poorest for alimentary tract medications (48%) [123]. Recall was influenced by the number of chronically used medications: 71% for one drug, 64% for two drugs, and 59% for three or more drugs, although duration of use was not related to recall. However, the questionnaire did not allow sufficient space to record all medications

used in the time period of this study. Thus, if respondents were unable to record all medications due to space limitations, a misleading validation might have occurred: it appeared that respondents were unable to recall all the medications dispensed according to the database.

Another methodologic study evaluated whether question structure influences the recall of currently used medications in 372 subjects with hypertension who had at least 90 days of dispensings in the PHARMO database [105]. The questionnaire had indication-specific questions first (e.g., medications used for hypertension, diabetes), followed by an open-ended question that asked if the subjects used any other medications not already mentioned. For hypertension, the sensitivity was 91% for indication-specific questions and 16.7% for open-ended questions. About 20% of subjects listed medications on the questionnaire that were not in the database, and a similar proportion failed to list medications on the questionnaire that were in use according to the pharmacy database. Based on these recall sensitivity results, indication-specific questions appear to invoke better recall accuracy. However, to adequately address the issue of question structure, a questionnaire could be designed to ask open-ended questions first, followed by indication-specific questions. This sequencing would allow a comparison of the number of medications recalled by each question structure.

Influence of Patient Population

Few studies have evaluated whether demographic and behavioral characteristics influence the recall of past medication use, but results to date suggest that recall does vary by these factors as well as by therapeutic class and study design. For example, research suggests that education attainment [104,108,124] and race/ethnicity [91,95] may affect recall accuracy. Studies are inconsistent for age [77,91,95–97,101,103,107,116], socioeconomic status [64,101,103,107,124], and smoking [95,97] as predictors of recall accuracy, and no study found that recall accuracy varies by gender [97,99,123]. The inconsistencies in the effect of age on recall accuracy might arise from differing study designs. The two studies that reported an age effect were methodologic studies evaluating recall accuracy [97,123], whereas the two that reported no age effects [91,95] were etiologic studies that reported verification of drug use as a measure of exposure misclassification for the association under study. Because of the paucity of information on predictors of recall, further research in this area is warranted.

Example

As indicated previously, accuracy of *de novo* questionnaire studies has been determined via comparison with pharmacy, general practitioner, and hospital records. To find an example of available study types, we conducted a literature scan of published studies, specifically searching for validation of NSAID use in questionnaire studies, and summarized our findings in Table 37.1.

Comparing use recalled during telephone interviews to a pharmacy database, West and colleagues found that 57% (95% CI 50–65%) of "any" NSAID use during the previous 12 years was accurately reported [97]. While a single dispensing was reported only 41% (95% CI 32–50%) of the time, repeated use was reported 85% (95% CI 76–94%) of the time, using the pharmacy records as the gold standard. Thirty percent of interviewees reported NSAID name and 15% reported both name and dose. Report was poorer with a shorter duration of use or over a longer recall period.

In summary, the methodologic literature on recall accuracy discussed above indicates that study participants have difficulty remembering drug use from the distant past, which contributes to misclassification of exposure in *de novo* studies. Researchers are using best practices in questionnaire design, including medication-specific and indication-specific questions, along with recall enhancements, which have been shown to produce better data. Calendars and photos of drugs augment recall to a greater degree than listing only the brand names of the

Table 37.1 Validation of NSAID exposure in studies using questionnaires.

Author	Recall period	Questionnaire and sample size	Study question	Memory aids	Comparison data source	Findings
West 1995 [97]	2–3 years 7–11 years	Telephone interviews n = 319	Nonsteroidal antiinflammatory drugs (NSAIDs)	Pictures of NSAIDs	Pharmacy database	Recall percentage for any NSAID use: 57 (95% CI 50–64) Single NSAID dispensed in 12 year period: 41 (95% CI 32–50) Repeated NSAID use: 85 (95% CI 76–94)
			For those with repeated NSAID use, a single NSAID was selected as the target drug for assessing name, dose, and dates of use			NSAID name: 30 (95% CI 24–36) NSAID name and dose: 15 (95% CI 10–20) Agreement: (a) ±6 months, (b) ±1 year, (c) ±2 years (a) (b) (c) First use 20 28 51 Last use 17 24 42 Duration 67 71 80
Smith 1999 [115]	Current use	Personal interview and medication inventory n = 55 users	Aspirin	None	Serum levels	0.16 (0.0–0.32)

drugs in question. These techniques – namely, photos, calendars, and the two different types of drug questions – have become the state of the art for collecting self-reported drug data by personal or telephone interview.

The literature to date suggests that recall accuracy of self-reported medication exposures is sometimes, but not always, influenced by type of medication, drug use patterns, design of the data collection materials, and respondent characteristics. Given the current state of the literature, epidemiologists who plan to use questionnaire data to investigate drug–disease associations will need to consider which factors may influence recall accuracy in the design of their research protocols.

Self-Reported Diagnosis and Hospitalizations from De Novo Studies

Accuracy and Recall

Just as recall accuracy of past medication use varies by the type of drug, the ability of respondents to remember disease conditions varies by disease, particularly when it is chronic, like hypertension, or is viewed as threatening, such as sexually transmitted infections. The best reporting has been

noted with conditions that are specific and familiar, such as diabetes mellitus [113,125–131], hypertension [113,126,128,129,132], asthma [125,127,128], and cancers such as breast, lung, large bowel, and prostate [129,132–134]. However, assessing reporting accuracy is likely more difficult for common, recurring symptom-based conditions, such as sinusitis, arthritis, low back pain, and migraine headaches, which many people may have, or believe they have, without having been diagnosed by a clinician. For recall of acute conditions such as fractures, there is typically good agreement between self-report and the comparison data source, although the one methodologic study of fracture incidence indicated a slight tendency for overreporting of hand, finger, rib, or facial fractures [135], which might be attributed to confusing a fracture with other similar orthopedic problems like sprains and strains. Recall of acute conditions is likely to depend on the length of the recall period: mild traumatic brain injury that occurred prior to age 10 years was poorly recalled 15 years later [136].

Three studies assessed the recall accuracy for self-reported mental illnesses, comparing respondent information to clinical evaluation [127,128,137]. The results indicated poor agreement between the two data sources, with underreporting as the primary reason for poor agreement. It is unclear from these studies whether the reason for underreporting was the respondent's unwillingness to admit to mental illness or whether the conditions were actually underdiagnosed.

Both underreporting and overreporting of diagnoses have been noted in studies comparing self-reported diagnoses to clinical records [127,128], with overreporting occurring for conditions in which the diagnostic criteria are less explicit [138]. For common ailments, underreporting was often the major cause of disagreement [113,125,129,131]. Both overreporting and underreporting were noted for cardiovascular conditions, depending on the data source used for comparison [113,126,128,129,131,132,134,139–141]. In most instances of recall error, many respondents who had incorrectly reported myocardial infarctions (MIs) and stroke had other conditions that they may have mistakenly understood as coronary heart disease, MI, or stroke, based upon communication with their physician during their diagnostic visits [134,139–141].

Influences on Accuracy

Influence of Questionnaire Design

Questionnaire design also influences validity of disease and hospitalization data obtained by self-report. Simpler questions yield better responses than more complex questions, presumably because complex questions require the respondent to first comprehend what is being asked and then provide an answer. Inherent redundancy in longer questions and allowing more time to develop an answer to the question may increase recall [142]. However, longer questions could increase the cost of the research and could needlessly tire the respondents, leading to satisficing. Facilitating recall by providing respondents with a checklist of reasons for visiting the doctor improves recall of all medical visits [143].

Although specific guidance on best practices for improving the ascertainment of diagnoses and hospitalizations is lacking, there are several general approaches to questionnaire design that are useful (see, for example, Sudman and Bradburn [144] and McColl and colleagues [145] for further details). Briefly, researchers developing questionnaires should be mindful of question wording and sequencing and response formats. With regard to question wording, to increase response accuracy, questionnaire designers should attend to the cognitive processes involved in developing a response, especially those related to saliency for the respondent. Whether a respondent recalls having been diagnosed with a particular condition previously is likely to depend on the seriousness of the condition. Use of a filter question such as, "Have you had any side effects from use of drug X in the past year?" must be done with caution because respondents who

avoid the filter are not asked subsequent questions that may be important for the study. As noted for recall of medications, open-ended questions are not recommended, particularly if the questionnaire is self-administered. That said, all potential response categories must be listed when using closed-ended questions or when an "other" category is provided. Questionnaire design experts suggest that demographic questions be placed at the end because they may be regarded as threatening.

The typical rule of thumb for question sequencing is to ask general questions before delving into specific topics and to group questions according to topic. When laying out the questions in a questionnaire, researchers should consider whether ordering effects are possible: for example, ask about heart disease in general before asking about a heart attack. Ordering might influence response rates to particular questions and may vary with the topic and make-up of the respondent population. With regard to response formats, the response categories should be unambiguous, nonoverlapping, and exhaustive. When there is a possibility of biased response due to response ordering, it is best to randomize the response options to minimize the bias. Finally, satisficing is also possible when respondents are asked to identify the diagnoses they have been given previously.

Influence of Patient Population

Factors influencing accuracy of past diagnoses and hospitalizations include the number of physician services for that condition and the recency of services [44–46,146–148]. For reporting of diagnoses, the longer the interval between the date of the last medical visit for the condition and the date of interview, the poorer the recall was for that condition [44–46]. These differences in recall may be explained in part by recall interval, patient age, a cohort (generational) effect, or some intertwining of all three factors. Diagnoses considered sensitive by one generation may not be considered as such by subsequent generations. Further, terminology changes over time, with prior generations using different nomenclature compared with recent generations.

Conditions with substantial impact on a person's life are more accurately reported than those with little or no impact. More patients with current restrictions on food or beverages due to medical problems reported chronic conditions that were confirmed in medical records than did those without these restrictions [44]. Similarly, those who had restrictions on work or housework reported their chronic conditions more often than those who did not have these restrictions [44]. The major determinant of recall for spontaneous abortions was the length of the pregnancy at the time the event occurred: nearly all respondents who experienced spontaneous abortions occurring more than 13 weeks into the pregnancy remembered them compared with just over half of respondents who experience such abortions occurring in the first six weeks of pregnancy.

Perhaps as a result of the emotional stress, lifestyle changes, and potential financial strain, hospitalizations tend to be reported accurately [147]. Further, underreporting of hospitalizations occurred in only 9% of patients who received surgery compared to 16% of patients without a surgical procedure. Underreporting in those with only a one-day hospital stay was 28% compared with 11% for 2–4-day stays and approximately 6% for stays lasting five or more days.

Surgical procedures are also more likely to be accurately recalled. General practitioner records confirmed 90% of the surgeries reported during one study interview. For the remaining 10%, the medical record may have lacked the needed information [149]. Recall of surgery date (±1 year) was correct for 87.5% of patients interviewed. Researchers also agree that respondents remember the type of surgery accurately [116,148–150]. Recall accuracy was very good for hysterectomy and appendectomy [110,125,129], most likely because these surgeries are both salient and

familiar to respondents. Cholecystectomy [129] and oophorectomy [110] were not as well recalled and were subject to some overreporting. However, overreporting may have been due to the potential incompleteness of the medical records used for comparison [110]. For induced abortions, marginal agreement occurred, as noted by records from a managed care organization: 19% of women underreported their abortion history, 35% overreported abortions, and 46% reported accurately according to their medical record [151].

The influence of demographic characteristics on reporting of chronic illnesses has been evaluated in many studies, although the results are conflicting. The most consistent finding is that overall recall accuracy decreases with age [113,116,131,133,152], although this may be confounded by recall interval or cohort (generational) effects. Whether gender influences recall accuracy is uncertain. Men have been reported to recall more accurately than women, independent of age [125], whereas conflicting evidence found that women reported more accurately than men [127], especially in older age groups [44]. Further studies indicate that gender and age differences depended upon the disease under investigation [127], with women overreporting malignancies and men overreporting stroke [131]. No differences were found for reporting of hospitalizations by age or gender [147].

Reporting of illnesses, procedures, and hospitalizations tends to differ by race/ethnicity, but most studies had much larger proportions of whites than nonwhites [44,116,125,127,147,151]. Reporting by education level was equivocal; one study showed no difference [46] while another study indicated better recall for those with less education [44], and others suggested more accurate responses for those with a college education [131,133,135,151]. Those with a poor or fair current health status reported conditions more completely than those with good to excellent health status [44].

Although menarche and menopause are not medical conditions *per se*, the age at which they occur is often of interest in pharmacoepidemiologic studies. In the Menstrual and Reproductive Health Study, which had recall periods ranging from 17 to 53 years (mean 33.9 years), the exact age of menarche was recalled by 59%, and age within one year was recalled by 90% [153]. Similarly, for menopause, 45% of women were able to report their exact age at natural menopause and 75.5% reported age within one year. The percentage agreements for surgical menopause were 55.6% and 83.4%, respectively, for exact age and age within one year. The lower percentage agreement for age at which natural menopause occurred compared to that for surgical menopause may be attributed to the gradual occurrence of natural menopause compared to the definitive nature of hysterectomy [154].

Example

We conducted a literature scan of published studies searching for outcomes of MI and GI bleeding associated with use of NSAIDs to provide specific examples of validation and reliability studies for diagnoses (Table 37.2). Many of those identified were methodologic studies conducted specifically to determine the accuracy of the questionnaire; however, some of the accuracy assessments were embedded in empirical studies. Fourrier-Reglat and colleagues compared reported medical data from patient and prescriber self-administered questionnaires [155]. Myocardial infarction showed substantial agreement (kappa = 0.75; 95% CI 0.71–0.80), while upper GI bleeding had only slight agreement (kappa = 0.16; 95% CI 0.11–0.22) between the two reporting groups. When the prescriber data were used as the gold standard, patient reports of MI provided moderately complete data (sensitivity 77.7%; specificity 99.6%; positive predictive value [PPV] 77.1%; negative predictive value [NPV] 99.6%), and reports of upper GI bleeding by patients were not typically confirmed by the prescriber reports (sensitivity 44.6%; PPV 10.4%).

Jarernsiripornkul and colleagues also used a multistage process to develop a questionnaire to

Table 37.2 Validation of myocardial infarction (MI) or gastrointestinal (GI) outcomes in patients with nonsteroidal antiinflammatory drugs (NSAIDs) in questionnaire data.

Author	Questionnaire and sample size	Study question	Comparison data source	Conditions	Findings
Ambegaonkar 2004 [192]	Gastrointestinal Toxicity Survey (NSAID Induced) (GITS [NI]) – 11 questions, n = 400 patients	To test a new questionnaire designed to identify patients at high risk for NSAID-associated GI events	Stanford Calculator of Risk for Events (SCORE) – 6 questions	56.0% rheumatoid arthritis	The overall correlation between results for GITS (NI) responses and the total score for the SCORE questionnaire was 0.96 (P<0.001) Comparison: ordinary least square R^2 = 0.91 feasible generalized least squares (FGLS) R^2 = 0.93 Use of the FGLS regression analysis and comparison of the risk levels predicted by the SCORE questionnaire and the GITS (NI) questionnaire demonstrated a 79.8% agreement for all four risk categories and an 88.8% agreement when the two highest risk categories were collapsed into a single category The multinomial logistic regression (MNL) analysis showed agreement of 75.8% for four risk categories and an agreement of 86.8% for three risk categories For both methods, disagreement was equally distributed among overprediction and underprediction of risk levels by the GITS (NI) questionnaire relative to the SCORE questionnaire. In the case of four risk categories, disagreement by two risk levels was limited to 0.6% and 1.5% for the FGLS and MNL regression methods, respectively

(Continued)

Table 37.2 (Continued)

Author	Questionnaire and sample size	Study question	Comparison data source	Conditions	Findings
Fourrier-Reglat 2010 [155]	CADEUS cohort (French national cohort study of traditional NSAIDs and COX-2 users conducted between September 2003 and August 2004 that employed self-administered questionnaires to obtain medical data from patients and their prescribers) n = 18 530 pairs of patients and prescribers	To compare patients and prescribers reported medical data	Prescribers report as gold standard		Previous medical history: MI: kappa = 0.75 (95% CI 0.71–0.80) Sensitivity: 77.7% Specificity: 99.6% PPV: 74.1% NPV: 99.6% Upper digestive hemorrhage: kappa = 0.16 (95% CI 0.11–0.22) Sensitivity: 44.6% Specificity: 98.5% PPV: 10.4% NPV: 99.8% NSAID indication: For index NSAID indication, the proportion of agreement ranged from 84.3% to 99.4% and concordance was almost perfect (kappa = 0.81–1.00) for inflammatory rheumatism, flu-like symptoms, dysmenorrhea and dental pain; substantial for arthritis, back pain and headache; moderate for osteoarticular pain.
Singh 1996 [193]	Stanford Health Assessment Questionnaire (HAQ)	To evaluate the event rates for all NSAID-induced GI complications in patients with rheumatoid arthritis, describe the time course of these events, and evaluate the role of prophylactic therapy with antacids and H2 receptor antagonists	Face validity and hospital records (2.4% hospitalized)		Face validity has been studied by surveying patients to ensure their understanding of the symptoms that are listed in lay language on the questionnaire; appropriate modification of the confusing symptoms has been made. Patient recall and accuracy in reporting side effects have been evaluated by repeat questionnaire administration, interview, and review of physician records. To minimize underreporting by patients, those events that are severe enough to require hospitalization are also ascertained by record review of all hospitalizations.

CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

have patients self-report potential adverse reactions to NSAIDs [156]. The questionnaire was designed to elicit adverse effects that would be reported to health professionals, to determine how well patient report compared with health professionals reporting to the Adverse Product Reaction Monitoring (APRM) Centre of the Thai FDA. The questionnaire was cognitively tested to finalize the version sent to the test sample along with pictures to facilitate recall. Of the 694 (42%) of questionnaires returned, 60% reported ≥1 symptom deemed as a possible or probably adverse drug reaction by a pharmacist. By comparison, only 5% of the self-reported symptoms indicative of adverse events from the questionnaires were recorded in the outpatient medical records.

These examples demonstrate the variation in methods used to collect and determine accuracy of questionnaire data. Although many methods are available for use, researchers should remember the principles discussed earlier in the chapter when they validate questionnaire data: not all validation is equivalent. Full disclosure of the process is important when reporting findings of any study.

In summary, the decision as to whether a person reports an illness during an interview appears to be related to age and type of illness, when it occurred, and its saliency, but is less likely to be mediated by demographic characteristics such as gender, race, and education. Illnesses that are considered embarrassing and that do not substantially alter the person's lifestyle are not reported completely, and these types of illnesses may change with each generation. Likewise, reporting accuracy depends on the consistency of documentation and the terminology utilized – from the questionnaire to the medical records – and finally, what has been communicated to the individual. Although difficult to measure, respondent motivation appears to influence the completeness of reporting as well [44,127,147].

Conducting Validation Studies to Assess Data Collected During Provision of Healthcare

In addition to conducting *de novo* studies to evaluate drug–disease associations, a variety of computerized, administrative claims, and EHR databases are available for pharmacoepidemiologic research, the structure, strengths, and limitations of which are reviewed in Chapters 14–14. One major advantage of using such databases for pharmacoepidemiologic research is the comparative validity of the drug data in lieu of questionnaire data, where recall bias is always a concern, as previously described.

In general, the administrative claims and EHR differ widely on many factors, such as size (e.g., from several hundred thousand to several million covered lives), number of health insurance plans and health systems included, the type of health services provided and therefore available for analysis (e.g., prescriptions, mental health benefits, general practice versus specialist visit data), inclusion of out-of-plan claims in the main database versus other databases, and the timeliness of the data (e.g., the lag for cleaning and obtaining data from a data vendor may be six or more months). The databases also differ on the number of available demographic variables: all have age (some may have date of birth) and sex, EHRs may have race (but administrative claims typically do not), or a measure of health status [157]. Because the administrative claims data were developed primarily for reimbursement, they all have relatively complete data on health service use and charges that are covered by the plan (and relatively incomplete data for services not covered by the plan). EHRs provide in-depth, granular data for a specific office or hospital visit but may not provide all health information for an individual in a longitudinal fashion, especially if the patient sees more than one healthcare provider. Linkage of EHRs and administrative claims can be resource intensive but may elucidate whether the data sufficiently capture the patient experience.

The drawbacks and limitations of these data systems are important to keep in mind. Their most critical limitation for pharmacoepidemiologic research is the manner in which health insurance is currently covered in the US, typically through the place of employment. If the employer changes plans, which may occur on an annual basis, or the employee changes among the plans offered by the employer, or the employee changes jobs, the plan no longer covers that employee or his or her family. In addition, the healthcare delivery system coverage for an employer may change over time. Thus, the continual enrolment and disenrolment of plan members hinder the opportunity for extended longitudinal analyses in both administrative claims and EHRs.

Best Practices for Validation Studies in Administrative Claims or EHR Databases

For the data in administrative claims or EHRs to be considered valid, people who appear in the computerized files as having a drug exposure or disease should truly have that attribute and those without the exposure or disease should truly not have the attribute. Validity and completeness are determined by comparing the database information with other data sources, such as comparison of paper medical records or EHRs, administrative claims, pharmacy dispensings, or procedure logs. Choice of an appropriate comparator varies by study question, variables used for the research study, the comparator, and availability of other data sources.

The study investigator must be aware of the limitations of both the administrative claims database and the chosen comparison dataset. The chosen comparator should provide sufficient data to validate both the exposure and outcome used for the study. A variable that provides linkage between the files in a data source, such as a medical record number, should be available so that accuracy can be evaluated within a subset of known study patients. For example, if a single claim contains six diagnosis codes and six months of claims were used to determine outcomes in patients, then all six diagnosis codes for all claims across the six-month study should be available in a comparison dataset to establish the validity of the outcome. As described earlier in the chapter, a validation assessment should include evaluation of patients with and without the exposure or outcome. Positive predictive value, negative predictive value, sensitivity, and specificity combined provide a complete picture of the agreement between the two data sources.

The following is a broad overview of how to conduct a validation study in administrative claims or EHR data. First, choose a meaningful number of patients for validation. This sample size should be statistically grounded; however, considerations of data availability, cost, and labor are understandable. Next, extract the variables needed to determine cohort selection, exposure, outcome, and other variables for validation. Calculate measures of agreement and error rates (e.g., standard deviations) between the two datasets. Finally, consider strengths and limitations of the two datasets to ascertain validity and completeness of the data source to answer the study question.

Influence of Data System

Completeness and validity of data are the most critical elements in the selection of a database for research. Completeness is defined as the proportion of all exposures, events of interest or both that occurred in the population covered by the database that appear in the computerized data. Missing subjects, exposures, or events could introduce bias in the study results [158]. For example, completeness of the drug data might vary by income level if persons with higher incomes and drug co-payments choose to obtain their medications at pharmacies not participating in a prescription plan, which is how pharmacy data are collected. Similarly, a bias may be introduced in the association

between a drug and a serious adverse drug reaction if hospitalizations for that adverse reaction are missing, for example if the researcher only has access to the outpatient clinic EHR database.

Influence of Clinical Coding Systems

Diagnoses, procedures, medications, and other therapeutics are included in administrative claims and EHR data through structured coding systems. Each coding system has its own ontology and is separated into specific codes, based on an established hierarchy. Further, the coding systems are updated periodically to reflect changes in the practice of healthcare as well as to incorporate new therapies and processes. Both codes and the general structure and hierarchy differ between coding systems. In many cases, a single code from a coding system is insufficient to define a variable and an *algorithm* needs to be developed. The algorithm may contain multiple codes, a required timing for codes, and/or a sequential process for determining the level(s) of the variable. It is likely that an algorithm developed in one coding system will require translation to be comparable to another coding system. Algorithms for each clinical concept should be developed and validated separately

The International Classification of Diseases (ICD) is the standard for classification of diseases for clinical and research purposes [159] and is used in many administrative claims systems, such as for billing purposes. The ICD is updated periodically, and adoption is asynchronous by country. For example, through the fall of 2015, most US administrative claims systems were using the clinical modification version of ICD-9, while many European administrative claims systems began using ICD-10 in the 1990s, and ICD-11 was released in 2018. As with the transition to ICD-10, we anticipate a staggered approach to implementation, with countries in Europe and Canada adopting the ICD-11 system (long) before the US. The ontology differs between the ICD systems, and codes have been mapped between ICD-9 and ICD-10 [160]. However, there is not a one-to-one correlation between codes; there are approximately 14 000 diagnosis codes in ICD-9 compared with approximately 70 000 in ICD-10 [161]. Mapping between codes can be used as a starting point to develop algorithms [162], but various techniques in mapping may yield different results [163]. As with other coding systems, any validation should be conducted separately between ICD systems.

Influence of Structured and Unstructured Healthcare Data in Computerized Databases Containing Administrative Claims or EHR data

In addition to the structured data in administrative claims and EHR, many components of healthcare are captured within clinician notes, images and descriptions of procedure results, and other unstructured data. The performance of an algorithm can be enhanced through use of this unstructured information in addition to the structured data from coding systems. Unstructured data can be converted into structured information (e.g., manually) for a specific project, or the algorithm can be modified to improve algorithm performance as cases are identified over time [164,165]. Liao and colleagues compared the performance of algorithms including unstructured data to detect coronary artery disease to algorithms using only structured elements in the same data source to assess validity in three chronic disease patient cohorts [166]. They found that inclusion of unstructured data increased sensitivity in all three cohorts, with the most improvement seen in the cohort where coronary artery disease prevalence was lowest. Note that while previous algorithms are sometimes used for comparison [165], patient charts (electronic or paper) are still often used as the reference standard for assessing validation [167].

Influence of Distributed Data Systems

Multiple health data sources may be included within a single study or for ongoing surveillance.

Simultaneous assessment of multiple data sources allows for better understanding of a larger population while also observing a diverse set of patients [168]. These multidatabase studies or distributed data systems may have differences in information collected, coding systems, language (e.g., across different countries), and even the underlying practice of medicine and overarching system of healthcare. Thus, even in the situation where distributed data systems use a common data model, careful consideration is warranted regarding how to assess validity of drugs, other therapeutics, diagnoses, procedures, and health-related events within each administrative claims data source contributing to the distributed data system. Whenever EHR data are utilized, differences across sites warrant assessment of validity within each health system to improve overall accuracy [167].

Validity of Drug and Other Medical Intervention Data in Administrative Claims or EHR Databases

Accuracy

Drug data in administrative claims databases are often not validated. Administrative claims data contain billing of a prescription that is dispensed (i.e., "filled") but do not contain information on the provider writing the prescription (or on the underlying condition the prescription is intended to treat). While prescriptions that are dispensed but unclaimed by patients should be removed from billing, they may remain within the administrative claims data. Furthermore, dispensing data cannot address drug ingestion or adherence, and over-the-counter medications are not typically included in the database at all. Thus, despite the widespread use of claims data to assess drug use, the data may not be accurate and validity should be tested, particularly when using a new drug exposure or database (e.g., some data sources may not contain a drug because it is not "on formulary" and thus is unavailable within the health system or allowed by the health insurer). Similarly, sensitivity analyses should be performed to determine the susceptibility of the results to possible misclassification, even within known data sources.

Unclaimed prescriptions, estimated to occur for approximately 2% of all prescriptions, present an adherence issue in administrative claims data [169]. For every 1000 new prescriptions, an average of 16.5 are unclaimed [170]. Antiinflammatory and antiinfective drugs tend to be the therapeutic class most often unclaimed [170–173]. Two-thirds of unclaimed prescriptions were for new prescriptions [171], and a similar proportion tended to be for nonessential medications [174]. Many unclaimed prescriptions were telephoned into the pharmacy [169,171], and the most frequent reason patients cited for not picking up a prescription was that they determined that they did not need the medication or they forgot to pick it up [169,172]. However, cost and having a similar medication at home were also often cited [169,172,174].

Drug data in EHRs represent the actual prescribing practice. EHR data account for the written prescription and may have sufficient detail to ascribe the prescription to the underlying condition it is intended to treat. Prescriptions or other documentation may also be available in the EHR for over-the-counter medications. However, they may not present a valid picture of the patient experience with the medication. The dispensing, ingestion, adherence, and pattern of use are typically not included as structured fields. Some of this information may be available in unstructured text such as the clinician visit notes.

Influences on Accuracy

Population and Representation in Data Source

One might ask how unclaimed prescriptions might affect the validity and completeness of pharmacy data. Many individuals have some type of pharmacy benefits plan in which reimbursement for medication costs is processed through a third-party payer. Entry into the reimbursement software is predicated on dispensing of the drug. However, a drug that is dispensed

but not claimed should be returned to stock and the appropriate adjustment be made to the patient's pharmacy benefits plan – failure to do this would be insurance fraud.

Unfortunately, when conducting research with pharmacy data, we do not know whether all such insurance adjustments have been made. So while we believe a substantial number of prescriptions were dispensed, they may not have been used at all. To the extent that dispensings in the database were not picked up, there is no chance that the individual had the drug exposure and our study would suffer from exposure misclassification. Exposure misclassification can occur even when dispensings were picked up but not actually used by patients. For these reasons, some researchers require a minimum of two dispensings for assessing patient exposure to chronic medications. This rule of thumb is thought to improve to the likelihood that the drug was taken by the patient.

Example

A handful of studies to date have assessed dispensing associated with prescriptions via linked administrative claims and EHR data. These studies indicated that 70–77% of initial prescriptions are dispensed [175,176]. Prescribed analgesics (i.e., pain medications, including NSAIDs) and lifestyle drugs (e.g., phosphodiesterase type 5 inhibitors) are least likely to be dispensed, while antimicrobials are most likely to be dispensed for an initial prescription. Substantial variation in dispensing was seen across medications within a class. In addition, results from Rowan suggest that <20% of patients taking analgesics and NSAIDs possessed adequate medication to be adherent throughout a 12-month period [176]; this finding may be consistent with intermittent or "as needed" utilization.

In summary, drug and medical intervention data are often considered to be correct when using administrative claims data and EHRs for research. Although this is generally the case, researchers should be aware of whether and how prescribing, dispensing, and administration of drugs are captured within each database they are contemplating using. We will likely see greater emphasis on data linkage and incorporation of more unstructured data from clinical notes into pharmacoepidemiologic research, which may lead to increased need for validation of drug and medical intervention exposures in the future.

Validity of Diagnosis, Procedure, and Hospitalizations in Administrative Claims and EHR Databases

Accuracy

Unlike the drug data, where many researchers are comfortable with data accuracy and completeness, inpatient and outpatient diagnoses in these databases raise considerable concern for investigators. The accuracy of outpatient diagnoses is more uncertain than inpatient diagnoses for several reasons. Hospitals employ experienced people to code diagnoses for reimbursement, which may not occur in individual physicians' offices where outpatient diagnoses are determined. Also, hospital personnel scrutinize inpatient diagnoses for errors [177], monitoring that does not typically occur in the outpatient setting.

Systematic errors as a result of diagnostic coding may influence the validity of both inpatient and outpatient diagnostic data. For example, diseases listed in administrative claims databases are often coded using the ICD coding system. Poorly defined diseases are difficult to code using the ICD system, and no way exists to indicate that an ICD code is coded for "rule-out" purposes. How healthcare plans deal with "rule-out" diagnoses is unclear; for example, should they be included or excluded from the diagnoses in the physician claims files? In a study of transdermal scopolamine and seizure occurrence, many patients with ICD codes indicating seizures had this diagnosis as a "rule-out" code when medical records were reviewed to confirm

the diagnosis, indicating that "rule-out" codes do become part of administrative claims data [178]. In addition, reimbursement standards and patient insurance coverage limitations may influence the selection of ICD codes for billing purposes [179]. The potential for abuse of diagnostic codes, especially outpatient codes, may occur when physicians apply to either an insurance carrier or the government for reimbursement and may be less likely in staff or group model health maintenance organizations (HMOs) such as Kaiser Permanente.

Influences on Accuracy

Validation Study Design

Abstraction of electronic data for validation studies is not subject to the issues of questionnaire design that are present with self-reported *de novo* studies; however, manual abstraction is subject to human error. Algorithms that are complex require substantial understanding of the healthcare environment in which the data were collected and necessitate review of lengthy portions of the patient chart, which may increase risk of error during record abstraction. Understanding of each specific healthcare system may be warranted to understand nuances of documentation practices. In one study, medical record documentation within a single multispecialty medical group showed that documentation varied across measures (e.g., medications documented 92% of the time, smoking history documented 38% of the time, and drug allergies documented in 62% of encounters) [180]. While no systematic patterns were noted across clinician and patient characteristics, differences in documentation were found between internists and pediatricians as well as between male and female providers.

Population and Representation in Data Source

At an institutional level, informaticists in the US have been concerned about the completeness of EHR data for research use [181,182]. Patient information in an EHR may be considered complete if it has sufficient detail regarding clinical encounters, if ongoing encounters are included over calendar time, if multiple types of data (e.g., labs, medications, and diagnoses) are available, and/or if sufficient information is available across a patient record to predict the condition of interest. In 2013, Weiskopf *et al.* reviewed all four of these definitions within a single healthcare system data warehouse in the United States and found that 26.9% of patients had complete records according to any one of the four definitions (8.4–18.5% of patient records were complete for each measure), and only 0.6% of patients had complete records according to all of these definitions of completeness [182].

Example

Continuing with the NSAID example, we conducted a literature scan of published studies validating MI or GI bleeding outcomes with use of NSAIDs in administrative claims databases; these studies are summarized in Table 37.3. Administrative claims data are often compared with medical records in a validation study. Most of these studies provide only a PPV that indicates whether the coding scheme is accurately classifying observed measures compared with another source. Validation measures such as sensitivity and specificity are not often calculated in these comparative studies.

In claims data, MI, denoted as ICD-9-CM code 410.xx, has been assessed in computerized health databases of Quebec [183], Saskatchewan Health [184], and the HealthCore® Integrated Research Database [185]. In all of these databases, this ICD-9-CM code had substantial or nearly perfect ability to validate the diagnosis of MI, with the PPV ranging from 88.4% to 96% across studies. Other ICD-9-CM codes used for possible detection of MI have shown poor ability to classify MI.

Both the overall PPV for ICD-9-CM 410.xx to measure MI and the PPV for MI among patients taking NSAIDs were evaluated in the HealthCore® Integrated Research Database. Among all the patients with a code for MI, the PPV was 88.4% (95% CI 83.2–92.5%). Among patients taking

Table 37.3 Validation of myocardial infarction (MI) and gastrointestinal (GI) bleeding events in studies using administrative claims data to evaluate harms of nonsteroidal antiinflammatory drug (NSAID) exposure.

Author	Dataset	Study aim and sample size	Comparison data source	Conditions	Findings
Abraham 2006 [188]	VA	To validate Veterans Affairs (VA) administrative claims data for the diagnosis of NSAID-related upper gastrointestinal events (UGIE) and to develop a diagnostic algorithm n = 906 ICD-9-CM codes and CPT procedure codes in patient treatment and outpatient care databases indicating upper gastrointestinal events (n = 606) Controls (n = 300)	Medical records	Case definition for UGIE was any of the following: Gastric ulcer 531.0, 531.1, 531.2, 531.3, 531.4, 531.5, 531.6, 531.7, 531.9 Duodenal ulcer 532.0. 532.1, 532.2, 532.3, 532.4, 532.5, 532.6, 532.7, 532.9 Peptic ulcer 533.0, 533.1, 533.2, 533.3, 533.4, 533.5, 533.6, 533.7, 533.9 Gastrojejunal ulcer with perforation 534.0, 534.1, 534.2, 534.3, 534.4, 534.5, 534.6, 534.7, 534.9 Gastrointestinal hemorrhage 578.0, 578.1, 578.9	Only ICD-9 codes for UGIE: Sens: 100% Spec: 96% PPV: 27% NPV: 100% ICD-9 and CPT for UGIE: Sens: 82% Spec: 100% PPV: 51% NPV: 99% ICD-9 and CPT algorithm for UGIE: Sens: 66% Spec: 88% PPV: 67% NPV: 88% Algorithm validated in additional 44 patients, PPV among NSAID users: 80%
Brophy 2007 [183]	Computerized health databases of Quebec, Canada	To determine whether a history of MI modified the risk of acute MI associated with the use of various NSAIDs n = 234 MI survivors	Previous validation of MI claims [194]; no validation of NSAID use	MI: hospitalization with ICD-9 code 410, considered fatal if person died within 30 days of admission	PPV = 0.96 (95% CI 0.94–0.98)

(*Continued*)

Table 37.3 (Continued)

Author	Dataset	Study aim and sample size	Comparison data source	Conditions	Findings
Castellsague 2009 [195]	Saskatchewan Health	To estimate the risk of upper gastrointestinal complications associated with use of cyclooxygenase-2 (COX-2) selective (celecoxib and rofecoxib) and individual nonselective NSAIDs compared with nonuse of these drugs Specific codes: n = 38 (10% sample) Nonspecific codes: n = 742 (all potential cases)	Medical records	Upper gastrointestinal complications: ICD-9 codes 531.0–531.2, 531.4–531.6, 532.0–532.2, 532.4–532.6, 533.0-533.2, 533.4–533.6, 534.0–534.2, 534.4–534.6, 569.3, 569.4, 569.8, 578	Previous research: PPV for site- and lesion-specific peptic ulcer disease codes in Saskatchewan = 91% PPV for nonspecific codes = 68% This study: Specific PPV = 92% Nonspecific code PPV (ranged across codes) = 60% for unspecified hemorrhage, 4% for hemorrhage of rectum/anus
Curtis 2008 [196]	Medicare	To evaluate the feasibility of adapting data mining methods using the empirical Bayes Multi-item Gamma Poisson Shrinkage (MGPS) algorithm to longitudinal administrative claims data Number not specified	Public use data files supplemented with specific medication data from CMS for greater precision in defining current NSAID exposure	Linked survey information, medical claims, and medication use data from the Medicare Current Beneficiary Survey (MCBS) for the years 1999–2003	"Identified current NSAID exposure using the MCBS medication data and all medical events using the linked Medicare claims"
van Staa 2009 [197]	GPRD	To evaluate the external validity of published cost-effectiveness studies by comparing the data used in these studies to observational data from actual clinical practice and whether these studies should have been used to inform prescribing policies. Selective COX-2 inhibitors (coxibs) and upper GI events were used as an example n = 96	Medical records	Upper GI events: ICD-10 codes K25–K29 NSAIDS: any prescription in GPRD	PPV = (95/96) = 99.0%

Varas-Lorenzo 2009 [184]	Saskatchewan Health	To evaluate risk of fatal and nonfatal acute MI with NSAID use n = 200	Medical records	ICD-9 code 410–414, 427.5, 798 ICD-10 code I20–I22, I23.3, I24–I25, I46, R96.0, R96.1, R98 Abstraction items included available information on cardiac symptoms; copies of available electrocardiograms recorded during the first 72 hours after hospital admission and the last one before hospital discharge; serum biomarkers levels: troponins, CPK-MB, or CPK measured within first 72 hours and compared with later measures; necropsy and other cardiac diagnostic test findings. Based on abstracted information, two cardiologists classified events as definite or probable/possible (either fatal or nonfatal) according to adapted standardized criteria recently adopted by American Heart Association/European Society of Cardiology. Classification of exposure to NSAIDs was based on the days between the index date and the end of supply of the most recent dispensing before the index date	PPV for ICD-9 code 410 = 0.95 (95% CI 0.91–0.98) PPV for ICD-9 code 411 for intermediate coronary syndrome = 0.73 (95% CI 0.70–0.77) PPV for ICD-9 code 411 for AMI = 0.09 (95% CI 0.07–0.11)

(*Continued*)

Table 37.3 (Continued)

Author	Dataset	Study aim and sample size	Comparison data source	Conditions	Findings
Wahl 2010 [185]	HealthCore® Integrated Research Database	To validate administrative claims codes with medical chart review for MI, ischemic stroke, and severe upper gastrointestinal (UGI) bleed events in a large, commercially insured US population n = 200 charts per outcome	Medical charts	MI: ICD-9 code 410.xx excluding 410.x2 and a length of stay (LOS) between 3 and 180 days, or death if LOS is <3 days Severe UGI bleed events were defined as a hospitalization for either UGI hemorrhage or peptic ulcer disease, including perforation. In the claims data, this was defined as ICD-9 codes 531.x, 532.x, 533.x, 534.x, 578.0, 578.1, 578.9, or a physician service code for GI hemorrhage (CPT code 43255 or ICD-9 procedure code 44.4x)	Overall: PPV for MI = 88.4% (177/200; 95% CI 83.2–92.5%) PPV for ischemic stroke = 87.4% (175/200; 95% CI 82.0–91.7%) PPV for severe UGI bleed = 56.5% (109/193; 95% CI 49.2–63.6%) Among those taking NSAIDs: PPV for MI = 92.3% (97/105; 95% CI 85.4–96.6%) PPV for ischemic stroke = 78.9% (57/72; 95% CI 67.6–87.7%) PPV for severe UGI bleed = 57.9% (70/121; 95% CI 48.5–66.8%)

NSAIDs, the PPV for MI was 92.3% (95% CI 85.4–96.6%). The difference between the overall PPV and PPV among patients taking NSAIDs could highlight the potential for differential coding by patient status. Further study of differences in diagnosis coding by medication or disease status is needed to know whether validating the drug and disease pair is warranted or whether validation of the exposure and outcome separately is sufficient to imply veracity of the results.

A substantial proportion of cases identified by algorithms for probable or definite MI within all databases are confirmed as probable or definite MI in medical records, with PPV ranging from 55% to 97%. Validity for MI has been measured in the Group Health Cooperative (now Kaiser Permanente Washington) [186] (sensitivity 86.5%; specificity 85.4%) and in the General Practice Research Database [187] (sensitivity 89.3%), with substantial agreement between the administrative claims and medical records.

Measurement of GI bleeding is more varied across databases, and several algorithms using different combinations of ICD-9-CM and CPT (Current Procedural Terminology) codes have been used to determine event occurrence. The PPV for the studies range from 60% to 100%. In general, the PPV has been higher when both ICD-9-CM and CPT codes are used. However, in the US Department of Veterans Affairs (VA) administrative claims data, where sensitivity and specificity were also assessed, the higher PPVs with use of both coding systems resulted in a lower sensitivity and specificity [188]. Limiting further to only those patients using NSAIDs, the PPV increased to 80%. Both the overall PPV for severe GI bleeding and the PPV for GI bleeding among patients taking NSAIDs were determined in the HealthCore® Integrated Research Database [185]. Among all patients with an ICD-9-CM or CPT code indicative of GI bleeding, the PPV was 56.5% (95% CI 49.2–63.6%). Among patients taking NSAIDs, the PPV for GI bleeding was 57.9% (95% CI 67.6–87.7%).

The variation seen in comparisons of GI bleeding in administrative claims and EHR may be due to the differences in algorithms used to determine GI bleeding. The variation may be due also to differences in GI bleeding in the underlying populations captured in each database. Validation, including measures of sensitivity and specificity, of the same algorithm in multiple databases will aid in determining whether GI bleeding can be adequately assessed in administrative claims and/or EHRs.

In summary, validating the case definition developed for observational studies using administrative claims databases with original documents such as inpatient or outpatient medical records is an important step to enhance the quality and credibility of the research. Although many studies in the past few years have reviewed original documents to validate the diagnoses under study or have referenced those validation studies, a need still exists for validation of drug exposures and disease diagnoses in databases in which no previous validation has been performed. As medical practice changes over time, further validation of previously validated claims is also warranted. Evaluating the completeness of the databases is much more difficult, as it requires an external data source that is known to be complete [143,187,189,190]. Although administrative claims and EHR databases have greatly expanded our ability to undertake pharmacoepidemiologic research, we need to ensure that our tools, including the databases used for our analyses, are complete and of the highest quality.

The Future

Methods for conducting pharmacoepidemiologic studies have shifted over the past several decades from reliance on studies requiring *de novo* data collection from individuals, to extensive use of electronic data from either administrative claims or EHRs, to linked data sources

and distributed data networks. Yet *de novo* data collection will continue to be required to ascertain information on quality of life, patient-reported outcomes (see Chapter 42), and medications either not included in pharmacy dispensing files or not reliably entered into EHRs, such as herbal and over-the-counter medications. In fact, with the advent of wearables and the Internet of Things, we anticipate that *de novo* collection of health data may increase in the coming years.

The improved computer technology that resulted in faster processor speeds and increased storage capacity facilitated storage of healthcare data in an electronic format, such as EHRs, and allowed development of distributed data networks using data from multiple health plans. The availability of these data for research has improved our ability to conduct studies [168] and the increasing uptake of EHRs is leading to increased availability of more granular clinical data for pharmacoepidemiologic research (e.g., lab results, clinical notes). Initial evaluation of EHR data suggests great promise, but increased data quality and standardization of terminology and codes will be required to make these data, collected for clinical care, useful for research purposes [191]. Similar processes will be warranted for use of data from wearables and prior to integration of new data from biobanks, mobile apps, social media, or other sources into a rigorous research framework.

As part of the standardization process, data holders will have to document that their data are valid for conducting research and surveillance activities. This will require investigators to apply their knowledge and practices from use of administrative claims and EHR data to linked data and to these novel data sources. Both medication exposure and outcome diagnosis data from these novel data sources do not carry the same level of comfort regarding validity as claims data and EHR data. As these data are considered for research, we hope and expect to see studies validating their use.

References

1 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**(1): 159–74.

2 Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol* 1993; **137**(11): 1251–8.

3 Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; **126**(2): 161–9.

4 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**(8476): 307–10.

5 Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. In: *Monographs in Epidemiology and Biostatistics*, vol. **10**. New York: Oxford University Press, 1986.

6 Jurek AM, Maldonado G, Greenland S, Church TR. Exposure–measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol* 2006; **21**: 871–6.

7 Chavance M, Dellatolas G, Lellouch J. Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study of the relation between handedness and immune disorders. *Int J Epidemiol* 1992; **21**(3): 537–46.

8 Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990; **132**(4): 746–8.

9 Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol* 1991; **134**: 1233–44.

10 Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 1992; **3**(3): 210–15.

11 Sorahan T, Gilthorpe M. Nondifferential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occup Environ Med* 1994; **51**: 839–40.

12 Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent differential misclassification. *Am J Epidemiol* 1991; **134**: 433–7.

13 Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol* 2005; **34**(3): 680–7.

14 Maldonado G, Greenland S, Phillips C. Approximately nondifferential exposure misclassification does not ensure bias toward the null. *Am J Epidemiol* 2000; **151**: S39.

15 Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008, pp. 144–5.

16 Greenland S. Multiple bias modelling for analysis of observational data [with discussion]. *J Roy Stat Soc Ser A* 2005; **168** (2): 267e306.

17 Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008.

18 Jurek AM, Maldonado G, Greenland S, Church TR. Uncertainty analysis: an example of its application to estimating a survey proportion. *J Epidemiol Commun Health* 2007; **61**: 650–4.

19 Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer, 2009.

20 Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014; **43**: 1969–85.

21 Leamer EE. Sensitivity analyses would help. *Am Econ Rev* 1985; **75**: 308–13.

22 Maldonado G. Adjusting a relative-risk estimate for study imperfections. *J Epidemiol Commun Health* 2008; **62**(7): 655e63.

23 Phillips C, Maldonado G. Using Monte Carlo methods to quantify the multiple sources of error in studies. *Am J Epidemiol* 1999; **149**: S17.

24 Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; **14**(4): 459–66.

25 Jurek AM, Maldonado G, Greenland S. Adjusting for outcome misclassification: the importance of accounting for case–control sampling and other forms of outcome-related selection. *Ann Epidemiol* 2013; **23**: 129–35.

26 Horwitz RI, Feinstein AR. Alternative analytic methods for case–control studies of estrogens and endometrial cancer. *N Engl J Med* 1978; **299**(20): 1089–94.

27 Hutchison GB, Rothman KJ. Correcting a bias? *N Engl J Med* 1978; **299**(20): 1129–30.

28 Greenland S. A mathematic analysis of the "epidemiologic necropsy". *Ann Epidemiol* 1991; **1**(6): 551–8.

29 Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**(4): 451–8.

30 Greenland S. Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Anal* 2001; **21**(4): 579–83.

31 Phillips CV, LaPole LM. Quantifying errors without random sampling. *BMC Med Res Methodol* 2003; **3**: 9.

32 Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003; **98**: 47–54.

33 Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004; **160**(4): 384–92.

34 Graff JJ, Sathiakumar N, Macaluso M, Maldonado G, Matthews R, Delzell E. The effect of uncertainty in exposure estimation on the exposure–response relation between 1,3–butadiene and leukemia. *Int J Environ Res Public Health* 2009; **6**: 2436–55.

35 Jurek AM, Lash TL, Maldonado G. Specifying exposure classification parameters for sensitivity analysis: family breast cancer history. *Clin Epidemiol* 2009; **1**: 109–17.

36 Jurek AM. Maldonado G. Quantitative bias analysis in an asthma study of rescue–recovery workers and volunteers from the 9/11 World Trade Center attacks. *Ann Epidemiol* 2016; **26**: 794–801.

37 Jurek AM, Maldonado G, Spector LG, Ross JA. Periconceptional maternal vitamin supplementation and childhood leukaemia: an uncertainty analysis. *J Epidemiol Commun Health* 2009; **63**: 168–72.

38 Maldonado G, Delzell E, Tyl RW, Sever LE. Occupational exposure to glycol ethers and human congenital malformations. *Int Arch Occup Environ Health* 2003; **76**: 405–23.

39 Scott LLF, Maldonado G. Quantifying and adjusting for disease misclassification due to loss to follow-up in historical cohort mortality studies. *Int J Environ Res Public Health* 2015; **12**: 12834–46.

40 Gordis L. Assuring the quality of questionnaire data in epidemiologic research. *Am J Epidemiol* 1979; **109**(1): 21–4.

41 Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge, MA: Cambridge University Press, 2000.

42 Berntsen D, Rubin DC. *Understanding Autobiographical Memory: Theories and Approaches*. Cambridge, UK: Cambridge University Press, 2012.

43 Kasabova A. *On Autobiographical Memory*. Newcastle upon Tyne: Cambridge Scholars, 2009.

44 Madow WG. Interview data on chronic conditions compared with information derived from medical records. *Vital Health Stat* 1967; **2**(23): 1–84.

45 Naleway AL, Belongia EA, Greenlee RT, Kieke BA Jr, Chen RT, Shay DK. Eczematous skin disease and recall of past diagnoses: implications for smallpox vaccination. *Ann Intern Med* 2003; **139**(1): 1–7.

46 Watson DL. Health interview responses compared with medical records. *Vital Health Stat* 1965; **1**(46): 1–74.

47 Gardner RS, Mainetti M, Ascoli GA. Older adults report moderately more detailed autobiographical memories. *Front Psychol* 2015; **6**: 631.

48 Warnecke RB, Sudman S, Johnson TP, O'Rourke D, Davis AM, Jobe JB. Cognitive aspects of recalling and reporting health-related events: Papanicolaou smears, clinical breast examinations, and mammograms. *Am J Epidemiol* 1997; **146**(11): 982–92.

49 Tourangeau R. Cognitive sciences and survey methods. In: Janine TB, ed. *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington, DC: National Academies Press, 1984, pp. 73–100.

50 Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. *Appl Cognit Psychol* 1991; **5**: 251–67.

51 Belli RF. The structure of autobiographical memory and the event history calendar: potential improvements in the quality of retrospective reports in surveys. *Memory* 1998; **6**(4): 383–406.

52 Geisen E, Murphy J. Compendium of web and mobile survey pretesting methods. International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2), Miami, FL, 2017.

53 Presser S, Couper MP, Lessler JT, *et al*. Methods for testing and evaluating survey questions. In: Presser S, ed. *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken: John Wiley & Sons, 2004.

54 World Bank. Living Standards Measurement Survey – Questionnaires. file:///C:/Users/Owner/Downloads/bihmane.pdf (accessed May 5, 2019).

55 Smith TW. Optimizing Questionnaire Design in Cross-National and Cross-Cultural Surveys. International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2), Miami, FL, 2016.

56 Q-Bank. https://wwwn.cdc.gov/qbank/ (accessed May 5, 2019).

57 Dean E, Caspar R, McAvinchey G, Reed L, Quiroz R. Developing a low-cost technique for parallel cross-cultural instrument development: the Question Appraisal System (QAS-04). *Int J Soc Res Methodol* 2007; **10**(3): 227–41.

58 Survey Quality Predictor (SQP) 2.1. http://sqp.upf.edu/ (accessed May 5, 2019).

59 Graesser AC, Wiemer-Hastings K, Kreuz R, Wiemer-Hastings P, Marquis K. QUAID: a questionnaire evaluation aid for survey methodologists. *Behav Res Methods Instrum Comput* 2000; **32**: 254–62.

60 University of Memphis. QUAID: Question Understanding Aid Tool. quaid.cohmetrix.com/ (accessed May 5, 2019).

61 Biemer PP, Lyberg L. *Introduction to Survey Quality*.Hoboken: Wiley Interscience, 2003, p. xiv.

62 Miller K. *Cognitive Interviewing Methodology*. Hoboken: Wiley, 2014.

63 Willis G. *Cognitive Interviewing*. Thousand Oaks: Sage Publications, 2005.

64 Willis GB. *Analysis of the Cognitive Interview in Questionnaire Design*. Oxford: Oxford University Press, 2015.

65 Krosnick JA. The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Survey Methods Newsletter* 2000; **20**: 4–8.

66 Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opin Quart* 1987; **51**: 201–19.

67 Geisen E, Bergstrom JR. Respondent–survey interaction. In: *Usability Testing for Survey Research*. Cambridge, MA: Morgan Kaufmann, 2017.

68 Nichols E, Anderson Riemer AE, Holland T, Olmsted Hawala EL. Best Practices of Usability Testing Online Questionnaires at the Census Bureau: How Rigorous and Repeatable Testing can Improve Online Questionnaire Design. International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2), Miami, FL, 2016.

69 Fowler S, Willis G. The Practice of Cognitive Interviewing Through Web Probing. International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2), Miami, FL, 2016.

70 Biemer P. Measurement errors in sample surveys. In: Pfeffermann D, Rao CR, eds. *Handbook of Statistics – Sample Surveys: Design, Methods and Applications*. Amsterdam: North-Holland, pp. 281–316.

71 Strom BL, Carson JL, Halpern AC, *et al*. A population-based study of Stevens–Johnson syndrome. Incidence and antecedent drug exposures. *Arch Dermatol* 1991; **127**(6): 831–8.

72 West SL, Ritchey ME, Poole C. Validity of Pharmacoepidemiologic drug and diagnosis data. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: John Wiley & Sons, 2012, pp. 757–94.

73 West SL, Strom BL, Poole C. Validity of pharmacoepidemiology drug and diagnosis data. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 3rd edn. Chichester: John Wiley & Sons, 2000, pp. 661–706.

74 West SL, Strom BL, Poole C. Validity of pharmacoepidemiology drug and diagnosis data. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 4th edn.

Chichester: John Wiley & Sons, 2000, pp. 709–766.

75 Lau HS, Florax C, Porsius AJ, de Boer A. The completeness of medication histories in hospital medical records of patients admitted to general internal medicine wards. *Br J Clin Pharmacol* 2000; **49**(6): 597–603.

76 Strom BL, Carson JL, Halpern AC, *et al*. Using a claims database to investigate drug-induced Stevens–Johnson syndrome. *Stat Med* 1991; **10**(4): 565–76.

77 Guess HA, West R, Strand LM, *et al*. Fatal upper gastrointestinal hemorrhage or perforation among users and nonusers of nonsteroidal anti-inflammatory drugs in Saskatchewan, Canada 1983. *J Clin Epidemiol* 1988; **41**(1): 35–45.

78 Kirking DM, Ammann MA, Harrington CA. Comparison of medical records and prescription claims files in documenting prescription medication therapy. *J Pharmacoepidemiol* 1996; **5**: 3–15.

79 Monson RA, Bond CA. The accuracy of the medical record as an index of outpatient drug therapy. *JAMA* 1978; **240**(20): 2182–4.

80 Christensen DB, Williams B, Goldberg HI, Martin DP, Engelberg R, LoGerfo JP. Comparison of prescription and medical records in reflecting patient antihypertensive drug therapy. *Ann Pharmacother* 1994; **28**(1): 99–104.

81 West SL, Strom BL, Freundlich B, Normand E, Koch G, Savitz DA. Completeness of prescription recording in outpatient medical records from a health maintenance organization. *J Clin Epidemiol* 1994; **47**(2): 165–71.

82 Buchsbaum DG, Boling P, Groh M. Residents' underdocumentation in elderly patients' records of prescriptions for benzodiazepine. *J Med Educ* 1987; **62**(5): 438–40.

83 Lain SJ, Roberts CL, Hadfield RM, Bell JC, Morris JM. How accurate is the reporting of obstetric haemorrhage in hospital discharge data? A validation study. *Aust N Z J Obstet Gynaecol* 2008; **48**(5): 481–4.

84 Crofts JF, Bartlett C, Ellis D, Fox R, Draycott TJ. Documentation of simulated shoulder dystocia: accurate and complete? *Br J Obstet Gynaecol* 2008; **115**(10): 1303–8.

85 Lefter LP, Walker SR, Dewhurst F, Turner RW. An audit of operative notes: facts and ways to improve. *ANZ J Surg* 2008; **78**(9): 800–2.

86 Haberman S, Rotas M, Perlman K, Feldman JG. Variations in compliance with documentation using computerized obstetric records. *Obstet Gynecol* 2007; **110**(1): 141–5.

87 Stange KC. The problem of fragmentation and the need for integrative solutions. *Ann Fam Med* 2009; **7**(2): 100–3.

88 West SL, Blake C, Zhiwen L, McKoy JN, Oertel MD, Carey TS. Reflections on the use of electronic health record data for clinical research. *Health Informat* 2009; **15**(2): 108–21.

89 Tsou AY, Lehmann CU, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy and paste in the EHR. Systematic review, recommendations, and novel model for health IT collaboration. *Appl Clin Inform* 2017; **8**(1): 12–34.

90 Glass R,Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**(4): 273–5.

91 Stolley PD, Tonascia JA, Sartwell PE, *et al*. Agreement rates between oral contraceptive users and prescribers in relation to drug use histories. *Am J Epidemiol* 1978; **107**(3): 226–35.

92 Adam SA, Sheaves JK, Wright NH, Mosser G, Harris RW, Vessey MP. A case–control study of the possible association between oral contraceptives and malignant melanoma. *Br J Cancer* 1981; **44**(1): 45–50.

93 Rosenberg MJ, Layde PM, Ory HW, Strauss LT, Rooks JB, Rubin GL. Agreement between women's histories of oral contraceptive use and physician records. *Int J Epidemiol* 1983; **12**(1): 84–7.

94 Persson I, Bergkvist L, Adami HO. Reliability of women's histories of climacteric oestrogen

treatment assessed by prescription forms. *Int J Epidemiol* 1987; **16**(2): 222–8.

95 Goodman MT, Nomura AM, Wilkens LR, Kolonel LN. Agreement between interview information and physician records on history of menopausal estrogen use. *Am J Epidemiol* 1990; **131**(5): 815–25.

96 Nischan P, Ebeling K, Thomas DB, Hirsch U. Comparison of recalled and validated oral contraceptive histories. *Am J Epidemiol* 1993; **138**(9): 697–703.

97 West SL, Savitz DA, Koch G, Strom BL, Guess HA, Hartzema A. Recall accuracy for prescription medications: self-report compared with database information. *Am J Epidemiol* 1995; **142**(10): 1103–12.

98 Strom BL, Schinnar R. An interview strategy was critical for obtaining valid information on the use of hormone replacement therapy. *J Clin Epidemiol* 2004; **57**(11): 1210–13.

99 Drieling RL, LaCroix AZ, Beresford SAA, Boudreau DM, Kooperberg C, Heckbert SR. Validity of self-reported medication use compared with pharmacy records in a cohort of older women: findings from the Women's Health Initiative. *Am J Epidemiol* 2016; **184**(3): 233–8.

100 Lewis JD, Brensinger C, Bilker WB, Strom BL. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf* 2002; **11**(3): 211–18.

101 Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health Nations Health* 1967; **57**(12): 2071–5.

102 Coulter A, Vessey M, McPherson K, Crossley B. The ability of women to recall their oral contraceptive histories. *Contraception* 1986; **33**(2): 127–37.

103 Feldman Y, Koren G, Mattice K, Shear H, Pellegrini E, Macleod SM. Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. *Teratology* 1989; **40**(1): 37–45.

104 West SL, Savitz DA, Koch G, *et al*. Demographics, health behaviors, and past drug use as predictors of recall accuracy for previous prescription medication use. *J Clin Epidemiol* 1997; **50**(8): 975–80.

105 Klungel OH, de Boer A, Paes AH, Herings RM, Seidell JC, Bakker A. Influence of question structure on the recall of self-reported drug use. *J Clin Epidemiol* 2000; **53**(3): 273–7.

106 Bryant HE, Visser N, Love EJ. Records, recall loss, and recall bias in pregnancy: a comparison of interview and medical records data of pregnant and postnatal women. *Am J Public Health* 1989; **79**(1): 78–80.

107 Mackenzie SG, Lippman A. An investigation of report bias in a case–control study of pregnancy outcome. *Am J Epidemiol* 1989; **129**(1): 65–75.

108 de Jong PCMP, Berns MPH, van Duynhoven YTHP, Nijdam WS, Eskes TKAB, Zielhuis GA. Recall of medication during pregnancy: validity and accuracy of an adjusted questionnaire. *Pharmacoepidemiol Drug Saf* 1995; **4**: 23–30.

109 Hulka BS, Kupper LL, Cassel JC, Efird RL, Burdette JA. Medication use and misuse: physician–patient discrepancies. *J Chronic Dis* 1975; **28**(1): 7–21.

110 Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**(1): 114–22.

111 Landry JA, Smyer MA, Tubman JG, Lago DJ, Roberts J, Simonson W. Validation of two methods of data collection of self-reported medicine use among the elderly. *Gerontologist* 1988; **28**(5): 672–6.

112 Johnson RE, Vollmer WM. Comparing sources of drug data about the elderly. *J Am Geriatr Soc* 1991; **39**(11): 1079–84.

113 Kehoe R, Wu SY, Leske MC, Chylack LT Jr. Comparing self-reported and

physician-reported medical history. *Am J Epidemiol* 1994; **139**(8): 813–18.

114 Law MG, Hurley SF, Carlin JB, Chondros P, Gardiner S, Kaldor JM. A comparison of patient interview data with pharmacy and medical records for patients with acquired immunodeficiency syndrome or human immunodeficiency virus infection. *J Clin Epidemiol* 1996; **49**(9): 997–1002.

115 Smith NL, Psaty BM, Heckbert SR, Tracy RP, Cornell ES. The reliability of medication inventory methods compared to serum levels of cardiovascular drugs in the elderly. *J Clin Epidemiol* 1999; **52**(2): 143–6.

116 Clegg LX, Potosky AL, Harlan LC, *et al*. Comparison of self-reported initial treatment with medical records: results from the prostate cancer outcomes study. *Am J Epidemiol* 2001; **154**(6): 582–7.

117 de Jong PCM, Berns M, van Duynhoven Y, *et al*. Accessibility and validity of data on medical drug use during pregnancy collected from various sources. *J Pharmacoepidemiol* 1991; **2**: 45–57.

118 Gama H, Correia S, Lunet N. Questionnaire design and the recall of pharmacological treatments: a systematic review. *Pharmacoepidemiol Drug Saf* 2009; **18**(3): 175–87.

119 Ademi Z, Turunen JH, Kauhanen J, Enlund H. A comparison of three questionnaire-based measures of analgesic use over 11 years in adult males: a retrospective analysis of data from a prospective, population-based cohort study. *Clin Ther* 2007; **29**(3): 529–34.

120 Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**(4): 670–6.

121 van Gelder MM, van Rooij IA, de Walle HE, Roeleveld N, Bakker MK. Maternal recall of prescription medication use during pregnancy using a paper-based questionnaire: a validation study in the Netherlands. *Drug Saf* 2013; **36**(1): 43–54.

122 Lunet N, Bastos J, Cumaio F, Silva P, Dias E, Barros H. Recall of drug utilization depends on subtle structural questionnaire characteristics. *Pharm World Sci* 2008; **30**(2): 175–81.

123 Van den Brandt PA, Petri H, Dorant E, Goldbohm RA, van de Crommert S. Comparison of questionnaire information and pharmacy data on drug use. *Pharm Weekbl* 1991; **13**(2): 91–6.

124 Cotterchio M, Kreiger N, Darlington G, Steingart A. Comparison of self-reported and physician-reported antidepressant medication use. *Ann Epidemiol* 1999; **9**(5): 283–9.

125 Linet MS, Harlow SD, McLaughlin JK, McCaffrey LD. A comparison of interview data and medical records for previous medical conditions and surgery. *J Clin Epidemiol* 1989; **42**(12): 1207–13.

126 Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health* 1989; **79**(11): 1554–6.

127 Commission on Chronic Illness. Chronic illness in a large city: the Baltimore Study. In: *Chronic Illness in the United States*, vol. **IV**. Cambridge, MA: Harvard University Press, 1957, pp. 297–328.

128 Heliovaara M, Aromaa A, Klaukka T, Knekt P, Joukamaa M, Impivaara O. Reliability and validity of interview data on chronic diseases. The Mini-Finland Health Survey. *J Clin Epidemiol* 1993; **46**(2): 181–91.

129 Spitz MR, Fueger JJ, Newell GR. The development of a comprehensive, institution-based patient risk evaluation program: II. Validity and reliability of questionnaire data. *Am J Prev Med* 1988; **4**(4): 188–93.

130 Midthjell K, Holmen J, Bjørndal A, Lund-Larsen G. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trondelag diabetes

study. *J Epidemiol Commun Health* 1992; **46**(5): 537–42.

131 Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ, Deeg DJ. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *J Clin Epidemiol* 1996; **49**(12): 1407–17.

132 Colditz GA, Martin P, Stampfer MJ, *et al*. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986; **123**(5): 894–900.

133 Bergmann MM, Calle EE, Mervis CA, Miracle-McMahil HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998; **147**(6): 556–62.

134 Paganini-Hill A, Chao A. Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. *Am J Epidemiol* 1993; **138**(2): 101–6.

135 Nevitt MC, Cummings SR, Browner WS, *et al*. The accuracy of self-report of fractures in elderly women: evidence from a prospective study. *Am J Epidemiol* 1992; **135**(5): 490–9.

136 McKinlay A, Horwood LJ. The accuracy of adult recall for early mild traumatic brain injury. *Disabil Rehabil* 2017; **39**(13): 1296–9.

137 Takayanagi Y, Spira AP, Roth KB, Gallo JJ, Eaton WW, Mojtabai R. Accuracy of reports of lifetime mental and physical disorders: results from the Baltimore Epidemiological Catchment Area Study. *JAMA Psychiatry* 2014; **71**(3): 273–80.

138 Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and medical records data: results from a case–control study on prostate cancer. *Int J Epidemiol* 1999; **28**(3): 409–17.

139 Tretli S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. *J Epidemiol Commun Health* 1982; **36**(4): 269–73.

140 Rosamond WD, Sprafka JM, McGovern PG, Nelson M, Luepker RV. Validation of self-reported history of acute myocardial infarction: experience of the Minnesota Heart Survey Registry. *Epidemiology* 1995; **6**(1): 67–9.

141 Walker MK, Whincup PH, Shaper AG, Lennon LT, Thomson AG. Validation of patient recall of doctor-diagnosed heart attack and stroke: a postal questionnaire and record review comparison. *Am J Epidemiol* 1998; **148**(4): 355–61.

142 Cannell CF, Marquis KH, Laurent A. A summary of studies of interviewing methodoloty. *Vital Health Stat* 1977; **1**(series 2): i–viii, 1–78.

143 Lessler JT, Harris BSH. *Medicaid Data as a Source for postmarketing Surveillance Information*. Research Triangle Park: Research Triangle Institute, 1984.

144 Sudman S, Bradburn N. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass, 2004.

145 McColl E, Jacoby A, Thomas L, *et al*. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess* 2001; **5**(31).

146 Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. *Am J Epidemiol* 1984; **120**(5): 727–33.

147 Biering-Sorensen U. Reporting of hospitalization in the Health Interview Survey. *Vital Health Stat* 1965; **1**(61): 1–71.

148 Corwin RG, Krober M, Roth HP. Patients' accuracy in reporting their past medical history: a study of 90 patients with peptic ulcer. *J Chronic Dis* 1971; **23**(12): 875–9.

149 Coulter A, McPherson K, Elliott S, Whiting B. Accuracy of recall of surgical histories: a

comparison of postal survey data and general practice records. *Community Med* 1985; **7**(3): 186–9.

150 Wingo PA, Ory HW, Layde PM, Lee NC. The evaluation of the data collection process for a multicenter, population-based, case–control design. *Am J Epidemiol* 1988; **128**(1): 206–17.

151 Udry JR, Gaughan M, Schwingl PJ, van den Berg BJ. A medical record linkage analysis of abortion underreporting. *Fam Plann Perspect* 1996; **28**(5): 228–31.

152 Linton KL, Klein BE, Klein R. The validity of self-reported and surrogate-reported cataract and age-related macular degeneration in the Beaver Dam Eye Study. *Am J Epidemiol* 1991; **134**(12): 1438–46.

153 Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. *Am J Epidemiol* 1979; **109**(2): 181–5.

154 Colditz GA, Colditz GA, Stampfer MJ, *et al*. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol* 1987; **126**(2): 319–25.

155 Fourrier-Reglat A, Cuong HM, Lassalle R, *et al*. Concordance between prescriber- and patient-reported previous medical history and NSAID indication in the CADEUS cohort. *Pharmacoepidemiol Drug Saf* 2010; **19**(5): 474–81.

156 Jarernsiripornkul N, Chaisrisawadsuk S, Chaiyakum A, Krska J. Patient self-reporting of potential adverse drug reactions to non-steroidal anti-inflammatory drugs in Thailand. *Pharm World Sci* 2009; **31**(5): 559–64.

157 Bonito AJ, Farrelly M, Han J, *et al*. *Assessment of the Feasibility of Creating a Managed Care Encounter-Level Database*. RTI Project Report 6703–003. Prepared for the Center for Organization and Delivery Studies. Rockville: Agency for Health Care Policy and Research, 1997.

158 Stergachis AS. Record linkage studies for postmarketing drug surveillance: data quality and validity considerations. *Drug Intell Clin Pharm* 1988; **22**(2): 157–61.

159 World Health Organization. Classification of Diseases (ICD 11). www.who.int/classifications/icd/en/ (accessed May 5, 2019).

160 Centers for Medicare and Medicaid Services. 2016 ICD-10-CM and GEMs. www.cms.gov/Medicare/Coding/ICD10/2016–ICD–10–CM–and–GEMs.html (accessed May 5, 2019).

161 National Center for Health Statistics. International Classification of Diseases, (ICD-10-CM/PCS) Transition – Background. www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm (accessed May 5, 2019).

162 Sun JW, Rogers JR, Her Q, *et al*. Adaptation and validation of the combined comorbidity score for ICD-10-CM. *Med Care* 2017; **55**: 1046–51.

163 Fung KW, Richesson R, Smerek M, *et al*. Preparing for the ICD-10-CM transition: automated methods for translating ICD codes in clinical phenotype definitions. *eGEMs* 2016; **4**(1): 4.

164 Liao KP, Cai T, Savova GK, *et al*. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ* 2015; **350**: h1885.

165 Upadhyaya SG, Murphree DH Jr, Ngufor CG, *et al*. Automated diabetes case identification using electronic health record data at a tertiary care facility. *Mayo Clin Proc Innov Qual Outcomes* 2017; **1**(1): 100–10.

166 Liao KP, Ananthakrishnan AN, Kumar V, *et al*. Methods to develop an electronic medical record phenotype algorithm to compare the risk of coronary artery disease across 3 chronic disease cohorts. *PLoS One* 2015; **10**(8): e0136651.

167 Newton KM, Peissig PL, Kho AN, *et al*. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc* 2013; **20**: e147–e154.

168 Trifiro G, Coloma PM, Rijnbeek PR, *et al*. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med* 2014; **275**: 551–61.

169 Schering Laboratories. *The Phantom Patient and Community Pharmacy Practice*. Schering Report SVII. Kenilworth: Schering, 1996.

170 Craghead RM, Wartski DM. An evaluative study of unclaimed prescriptions. *Hosp Pharm* 1991; **26**(7): 616–17, 632.

171 Farmer KC, Gumbhir AK. Unclaimed prescriptions: an overlooked opportunity. *Am Pharm* 1992; **NS32**(10): 55–9.

172 Hamilton WR, Hopkins UK. Survey on unclaimed prescriptions in a community pharmacy. *J Am Pharm Assoc* 1997; **NS37**(3): 341–5.

173 Kirking MH, Kirking DM. Evaluation of unclaimed prescriptions in an ambulatory care pharmacy. *Hosp Pharm* 1993; **28**(2): 90–1, 94, 102.

174 Kinnaird D, Cox T, Wilson JP. Unclaimed prescriptions in a clinic with computerized prescriber order entry. *Am J Health Syst Pharm* 2003; **60**(14): 1468–70.

175 Fischer MA, Stedman MR, Lii J, *et al*. Primary medication non-adherence: analysis of 195,930 electronic prescriptions. *J Gen Intern Med* 2010; **25**(4): 284–90.

176 Rowan CG, Flory J, Gerhard T, *et al*. Agreement and validity of electronic health record prescribing data relative to pharmacy claims data: a validation study from a US electronic health record database. *Pharmacoepidemiol Drug Saf* 2017; **26**(8): 963–72.

177 Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: strengths and limitations. *J Clin Epidemiol* 1989; **42**(10): 937–45.

178 Strom BL, Carson JL, Schinnar R, Snyden ES, Shaw M, Waiter SL. No causal relationship between transdermal scopolamine and seizures: methodologic lessons for pharmacoepidemiology. *Clin Pharmacol Ther* 1991; **50**(1): 107–13.

179 Wynia MK, Cummins DS, VanGeest JB, Wilson IB. Physician manipulation of reimbursement rules for patients: between a rock and a hard place. *JAMA* 2000; **283**(14): 1858–65.

180 Soto CM, Kleinman KP, Simon SR. Quality and correlates of medical record documentation in the ambulatory care setting. *BMC Health Serv Res* 2002; **2**(1): 22.

181 Kahn MG, Callahan TJ, Barnard J, *et al*. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *eGEMs* 2016; **4**(1): 18.

182 Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Informat* 2013; **46**: 830–6.

183 Brophy JM, Levesque LE, Zhang B. The coronary risk of cyclo-oxygenase-2 inhibitors in patients with a previous myocardial infarction. *Heart* 2007; **93**(2): 189–94.

184 Varas-Lorenzo C, Castellsague J, Stang MR, Perez-Gutthann S, Aquado J, Rodriguez LA. The use of selective cyclooxygenase-2 inhibitors and the risk of acute myocardial infarction in Saskatchewan, Canada. *Pharmacoepidemiol Drug Saf* 2009; **18**(11): 1016–25.

185 Wahl PM, Rodgers K, Schneeweiss S, *et al*. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 596–603.

186 Newton KM, Wagner EH, Ramsey SD, *et al*. The use of automated data to identify complications and comorbidities of diabetes: a validation study. *J Clin Epidemiol* 1999; **52**(3): 199–207.

187 Van Staa TP, Abenhaim L. The quality of information recorded on a UK database of primary care records: a study of hospitalizations due to hypoglycemia and other conditions. *Pharmacoepidemiol Drug Saf* 1994; **3**: 15–21.

188 Abraham NS, Cohen DC, Rivers B, Richardson P. Validation of administrative data used for the diagnosis of upper gastrointestinal events following nonsteroidal anti-inflammatory drug prescription. *Aliment Pharmacol Ther* 2006; **24**(2): 299–306.

189 Stergachis AS. Evaluating the quality of linked automated databases for use in pharmacoepidemiology. In: Hartzema AG, Porta MS, Tilson HH, eds. *Pharmacopeidemiology: An Introduction*. Cincinnati: Harvey Whitney, 1991, pp. 222–34.

190 Harris BL, Stergachis A, Ried LD. The effect of drug co-payments on utilization and cost of pharmaceuticals in a health maintenance organization. *Med Care* 1990; **28**(10): 907–17.

191 Kudyakov R, Bowen J, Ewen E, *et al*. Use of an electronic health record to classify patients with newly diagnosed versus pre-existing type 2 diabetes: building methodology infrastructure for comparative effectiveness research. *Population Health Manage* 2012; **15**(1): 3–11.

192 Ambegaonkar A, Livengood K, Craig T, Day D. Predicting the risk for gastrointestinal toxicity in patients taking NSAIDs: the Gastrointestinal Toxicity Survey. *Adv Ther* 2004; **21**(5): 288–300.

193 Singh G, Ramey DR, Morfeld D, Shi H, Hatoum HT, Fries JF. Gastrointestinal tract complications of nonsteroidal anti-inflammatory drug treatment in rheumatoid arthritis. A prospective observational cohort study. *Arch Intern Med* 1996; **156**(14): 1530–6.

194 Levy AR, Tamblyn RM, Fitchett D, McLeod PJ, Hanley JA. Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction. *Can J Cardiol* 1999; **15**(11): 1277–82.

195 Castellsague J, Holick CN, Hoffman CC, Gimeno V, Stang MR, Perez-Gutthann S. Risk of upper gastrointestinal complications associated with cyclooxygenase-2 selective and nonselective nonsteroidal antiinflammatory drugs. *Pharmacotherapy* 2009; **29**(12): 1397–407.

196 Curtis JR, Cheng H, Delzell E, *et al*. Adaptation of Bayesian data mining algorithms to longitudinal claims data: coxib safety as an example. *Med Care* 2008; **46**(9): 969–75.

197 van Staa TP, Leufkens HG, Zhang B, Smeeth L. A comparison of cost effectiveness using data from randomized trials or actual clinical practice: selective cox-2 inhibitors as an example. *PLoS Med* 2009; **6**(12): e1000194.

38

Studies of Medication Adherence

Julie Lauffenburger[1], Trisha Acri[2], and Robert Gross[3]

[1] *Division of Pharmacoepidemiology and Pharmacoeconomics and Center for Healthcare Delivery Sciences, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA*
[2] *Philadelphia, PA, USA*
[3] *Departments of Medicine (ID) and Epidemiology/Biostatistics/Informatics, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*

In this chapter, we will describe the importance of adherence in pharmacoepidemiologic research, the methods for measuring adherence, methodologic issues that arise once adherence has been measured, and future directions. While we use many different drug–disease examples, we focus on examples from HIV and cardiovascular diseases because these areas have been major focuses of adherence research.

The underuse of essential medications imposes significant clinical and financial burdens on healthcare systems. Data show that as many as half of patients do not take their medications as prescribed, resulting in more than $100 billion in excess annual spending in the US [1]. Nonadherence is also thought to contribute to 11% of US hospitalizations each year [1]. Without accurate measurements of adherence incorporated into research and practice, the problem will remain underappreciated and poorly addressed.

Despite its importance, medication adherence is difficult to define. Earlier research has used the term *compliance*, or "the extent to which the patient's dosing history conforms to the prescribed regimen," to describe this behavior [2]. However, this term implies that patients passively "conform" to the prescriber's directions; therefore, the term *adherence* is now strongly preferred [3]. *Adherence* better conveys the idea of a patient–provider relationship where the patient implements the provider's recommendations.

Another reason why adherence has been difficult to define is that it is not a single static behavior but instead encompasses a set of behaviors over time. One common taxonomy developed by a scientific consensus group classifies adherence along three phases: (1) initiation, (2) implementation, and (3) persistence (Figure 38.1) [4]. *Initiation* describes initial engagement with the prescribed medication. Research suggests that as many as 30% of newly prescribed therapies are never actually filled by patients [5], which is often referred to as primary nonadherence. *Implementation* represents how well the patient follows the prescribed regimen while s/he is engaged with treatment. While varying greatly across diseases, approximately 50% of patients are thought to not correctly follow prescribed regimens. *Persistence* refers to how long the patient continues to follow the regimen [6].

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

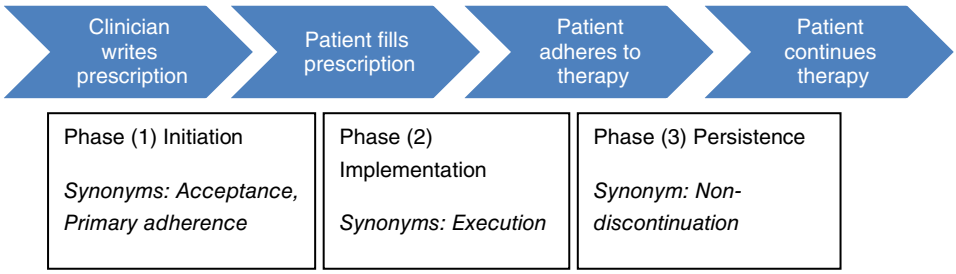


Figure 38.1 Phases and taxonomy of adherence.

Poor treatment adherence can occur along any of these phases.

The actual behaviors involved in taking a prescribed medication as directed become more complicated when considering each adherence phase. This taxonomy helps distinguish patients who never initially fill a prescription from patients who occasionally forget to take doses as well as patients who take a medication regularly at first but then later discontinue. Classifying all three types of patients simply as "nonadherent" ignores the fact that each of these patients may differ with respect to treatment outcomes and likely have different adherence barriers requiring different interventions [3]. This taxonomy also highlights the multifactorial behaviors required for sustained medication adherence and why measuring it, identifying the barriers, and then improving adherence has been difficult. Regardless, practical approaches to measuring and analyzing adherence have been successfully developed, and we will discuss the challenges and utility of various approaches to measuring adherence throughout the chapter.

Clinical Problems Addressed by Pharmacoepidemiologic Research

Adherence research confronts the truism attributed to former US Surgeon General C. Everett Coop, MD that "drugs don't work in patients who don't take them" [7]. Measuring adherence is essential in order to address several issues in the interpretation of studies of beneficial and adverse effects of medications. In randomized trials, treatment adherence can be an important factor that affects the estimates of efficacy and safety of the tested medications (see Chapter 32). Poor adherence to the drug being tested can lead to underestimates of drug efficacy [8]. Further, information about adherence allows for a more accurate assessment of drug safety because those who do not take the drug cannot experience its toxicity. Because perfect adherence is not attainable, even in clinical trials, measurements of adherence can elucidate whether a drug fails to exert an effect because it did not work or because it was not taken properly. Poor adherence may itself also be a marker of toxicity or adverse events.

Once a medication is marketed, information from clinical trials gives only a limited view of how drugs are used by patients. Patients who volunteer for clinical trials are thought to be more motivated than those in usual care [8,9]. Therefore, measuring adherence in observational studies of drug effectiveness and safety may be even more important than in clinical trials. Furthermore, assessing adherence in observational studies provides a more "real-world" estimate of adherence in clinical populations. Finally, because adherence itself is a major determinant of treatment outcomes, it can also be the specific focus of pharmacoepidemiologic research.

Nonadherence can be intentional or unintentional. Studies have identified many potential barriers to adherence, broadly categorized as patient-, system-, and medication-specific factors. Common patient barriers consist of forgetting to take the medication, lack of knowledge or health literacy, and psychosocial factors such as depression and lack of social support [10–12]. System barriers include logistical difficulty in obtaining the medication and, in some settings, sporadic drug unavailability ("stock outs") [13,14]. Key medication-specific factors include regimen complexity and adverse effects [15,16]. Further, patients may decide on a dose-by-dose basis whether to take medicine as prescribed, perhaps to avoid side effects at inconvenient times (like avoiding increased urination at night). Finally, postmarketing studies have observed "pill fatigue" (in that adherence can decrease over time from being emotionally overwhelmed by taking medication), particularly when patients are followed for longer than typically done in trials [17,18]. It is a well-known phenomenon that the optimal adherence seen early in therapy often decreases over time [19]. Thus, observational adherence studies provide unique data not available from trials.

While missed doses are a more common adherence problem, taking extra doses can also be a problem. For example, extra doses of drugs with a narrow therapeutic window, such as warfarin for anticoagulation, may result in toxicity [20]. Patients may also take extra doses of narcotics prescribed for the treatment of pain because of inadequate pain relief or for potential abuse (see Chapter 28).

Measuring adherence can also be useful for determining the threshold of how much medication must be taken to obtain desired clinical outcomes; these dosing thresholds likely differ by drug and disease. In hypertension, taking at least 80% of prescribed medication has been an acceptable standard for blood pressure control [21]. However, in HIV, 80% adherence is often insufficient. For example, in a study of patients starting protease inhibitors for HIV, those who took 80–95% of doses were more likely than those with lower adherence to achieve complete suppression of viral replication [22]. Unfortunately, such detailed information is not available for most drugs and diseases. Despite the likelihood that 80% of doses taken is not the optimal universal cut-point for acceptable adherence, this threshold persists across research and quality measures [23]. Therefore, the default adherence goal should be to encourage the patient to take as many prescribed doses as possible, and future research is focusing on identifying more empiric and robust dose–response thresholds for various diseases.

Finally, adherence can also impact public health, especially in infectious diseases. For example, in tuberculosis and HIV, nonadherence can actually lead to resistance to medications. Because these resistant diseases are transmissible [24], the measurement of nonadherence and adherence interventions takes on greater public health importance.

Methodologic Problems to be Solved by Pharmacoepidemiologic Research

Challenges in the Measurement of Adherence

The gold standard for measuring adherence to treatments is directly observed therapy [14]. However, this approach is only practical in limited settings, such as the administration of a novel agent in a controlled environment. While many approaches exist, as will be discussed later, whatever the approach, the discovery of nonadherence in clinical settings can be embarrassing for patients, because it can imply lack of respect for the provider's advice or for one's welfare. Thus, knowledge that one's adherence is being monitored risks influencing the behavior it is

measuring (i.e., a Hawthorne effect). Moreover, tracking a daily activity can be burdensome regardless of whether individuals are aware of their own nonadherence. Therefore, measuring adherence requires creative approaches to accurately capture a daily activity performed at different times per day for different individuals.

Challenges in the Analysis of Adherence Data

Once adherence is measured, there are various approaches to analyzing the data depending on the data sources used. In clinical trials, adjusting results for adherence is complicated by the fact that being adherent itself is associated with better outcomes (i.e., placebo effect). For example, in a randomized double-blind placebo-controlled trial of propranolol after myocardial infarction, poor adherers had a 2.8 higher odds of mortality compared with good adherers in the same active arm, after adjustment. However, the adjusted odds ratio of mortality in those with poor adherence to placebo was, similarly, 2.7 [25]. Presumably, adherence to either agent, whether propranolol or placebo, was strongly associated with other unmeasurable lifestyle factors associated with mortality. How to control for this healthy adherer effect is an important analytic consideration.

Other analytic challenges include the duration and timing of adherence measurements. Because adherence behaviors vary over time, individuals may have substantial changes across the observational period. For example, individuals are prescribed lifelong regimens for many chronic diseases. When initiating treatment, adherence over the first 12 weeks may not be the same as adherence over the final 12 weeks. Simply summing adherence over an entire 52-week interval will provide an average of adherence, but short periods of nonadherence can substantially impact clinical outcomes [26]. Therefore, when conducting adherence analyses, researchers need to carefully consider the appropriateness of the adherence "interval(s)". Many adherence studies of chronic medications choose intervals that are at least 180 or 365 days long to capture enough variation in use; however, this choice must be balanced with the length of follow-up available on patients to ensure external generalizability [27].

Whatever the interval, the summation of adherence data can also be accomplished in different ways. The simplest is the percent of doses taken, but this may not be the most clinically relevant metric. Depending on drug pharmacokinetics and pharmacodynamics, gaps and variability in adherence may be more important than the proportion of prescribed doses taken. However, a composite measure of percentage of doses taken over an entire time period is often still used as the sole adherence measurement in research publications and measures of healthcare quality. There have been recent advancements in measuring adherence, which are discussed later in this chapter.

Additionally, many diseases are treated with combination therapy (either multiple medications in the same formulation or multiple separate formulations). When drugs are studied in combination to determine their effect (e.g., antihypertensive or antituberculous therapies), it is challenging to determine how to weight differential adherence or switching among the drugs [27]. Many of these issues can be addressed with currently available solutions, although methodologic challenges remain to be solved.

Currently Available Solutions

There are many different methods for measuring medication adherence, and each method has strengths and weaknesses. Which method is most appropriate depends upon the situation in which it will be used and how precise the measurement needs to be. Some measurements require more intensive patient-level contact than others, and some provide more granular

data with respect to timing of dose taking. For example, in prospective clinical trials, because of the direct patient contact, many of these techniques can be used. In other settings like retrospective studies using databases, options are more limited. Therefore, the use of multiple measures or sources of data may be helpful to confirm findings. For all approaches, the interpretation of adherence findings may also change depending on whether incident users or prevalent users of medication are examined, as adherence tends to be higher among prevalent users, in part because discontinuation is highest in the first few months after initiation. Therefore, many studies focus on incident users, but there are situations in which studying prevalent users may be more relevant, especially because new initiators are only a small proportion of all patients using a therapy at a given time [28–30].

We will describe each of the strategies, their strengths and weaknesses, and discuss considerations for the timing of assessing adherence.

Specific Techniques for Measuring Adherence

Self-reports

Among many approaches to assessing adherence, patient self-reported measures asking respondents about their adherence behaviors have been the most common method. They are simple, relatively inexpensive, quick and feasible, and can be obtained over the telephone, in person, or with paper or electronic surveys. Self-reported measures vary greatly in the phrasing of their questions, recall periods, and response items. Several different validated methods for assessing self-reported adherence are described here.

Self-reported adherence measures range from one-item questions inquiring about the frequency of missed doses to longer multi-item assessments evaluating beliefs associated with adherence and identifying barriers to adherence [31]. Most self-reported measures involve count or estimation-based recall focused on the implementation phase, in which respondents report the number of doses missed or taken within an interval or to estimate their overall execution of adherence. Some scales use a recommended adherence cutpoint while other scales identify a continuous measure of the degree of adherence.

In a systematic review, Nguyen *et al*. identified 43 validated self-reported adherence scales in the English language [32]. Perhaps the most common self-reported adherence tool historically used is the eight-item Morisky Medication Adherence Questionnaire (MMAS-8) [33,34]. However, the use of this scale requires licensing fees. The adult AIDS Clinical Trials Group (ACTG) adherence questionnaire [12] and Brief Medication Questionnaire [35] are other examples of common, publicly available tools that explore both behaviors and barriers to adherence. A recent three-item tool by Wilson *et al*. queries patients about how many days they missed medications over the last 30 days [36]. Other studies have used a single measure such as a visual analog scale, which asks participants to mark a point on a line from 0% to 100% to indicate the amount of medication taken over a specified recent time period [37]. Overall, the choice of measure may depend on the context of its use (e.g., clinical use or research), the burden to patients, and the disease states in which it has been validated. In addition, self-report may be the easiest method for clinicians to administer and more easily used to isolate the reasons for poor adherence for targeting interventions.

Self-reported adherence measures are moderately correlated with methods using electronic drug monitoring (EDM) or pharmacy dispensing data (described later in the chapter), though concordance can vary depending on the patient's level of adherence or the measurement window [31,38–40]. For example, in a study comparing three-day, seven-day, and one-month self-reports, the one-month window best approximated adherence obtained using EDM [41,42]. On the other hand, because of potential

overreporting, self-reported measures are thought to have high specificity and low sensitivity (i.e., self-reported nonadherence is generally accurate, while high self-reported adherence may not be accurate) [31]. Either way, self-reported measures have shown weaker associations with clinical outcomes than EDM or dispensing data [31,42].

There are some additional limitations to self-reported adherence measurements. Though they can be self-administered in high-literate patients or conducted by an interviewer, they are all limited by a patient's ability to recall missed doses and may be subject to social desirability bias (i.e., overreporting adherence to please providers or researchers). Social desirability can be mitigated by acknowledging the difficulty of always taking all medications. Interviews are also potentially limited by language barriers, poor literacy, time burden, and difficulty with medication names. Using computer-assisted self-administered interview can reduce these barriers by reading instructions and questions aloud and including high-resolution photographs of the medicines. These questions can be administered at a kiosk or computer in a waiting room. Empirical data suggest that computer-aided self-reports are less likely to overestimate adherence [43]. However, poor patient recall is still a problem, and self-reported measures are also limited by their ability to precisely describe the timing or patterns of dose taking.

Pharmacy Dispensing Data

Pharmacy dispening measurement was pioneered in the late 1980s and has been widely used in various chronic diseases [44]. These measures typically derive from secondary data from health insurers and are some of the most common ways of measuring adherence in pharmacoepidemiology.

Pharmacy dispensing data are generally considered to be accurate because the dispenser (e.g., a pharmacy) would not get reimbursed by insurance if the medication fills are not recorded. Compared with self-reported data, pharmacy dispensing data are not biased by poor recall, can be obtained from computerized records, and can be assessed retrospectively [38]. Another advantage is that the data can be easily processed by software and are available on large numbers of patients (often millions in the same database). However, the data quality may be less assured in settings where such tracking is less crucial for reimbursement or if prescriptions are obtained outside insurance plans [45]. Some approaches use data from pharmacies directly to capture all medications dispensed to patients and not just those paid by insurers. Also, in the US, these pharmacy dispensing data are generally only accurate for medications dispensed in the outpatient setting, because medications are not specifically paid for separately during hospitalization [46]. For questions related to adherence to one-time prescriptions (e.g., short courses of antibiotics), these data may not be useful beyond studying primary nonadherence because repeat dispensings are required to calculate an amount of medication consumed.

There are several different methods for measuring adherence using pharmacy dispensing data. In all approaches, adherence is measured indirectly based on patterns of medication dispensings (using the dispensing date and days supplied) by generating a "drug supply diary" that strings together consecutive medication dispensings based on the dates on which medications are dispensed to the patient at the pharmacy and the duration of the supply dispensed [19,47]. This supply diary can adjust for overlapping fills (e.g., truncating the days supplied for medications which are refilled before the medication supply from the prior dispensing would have been exhausted) and any known interruptions that may have occurred (e.g., by hospitalization). When generating the supply diary, researchers generally consider medications that are chemically related and not intended for use in combination to be interchangeable (e.g., two beta-blockers). For example, patients may

initiate one beta-blocker and later switch to a different beta-blocker. In this case, beta-blocker adherence is often measured continuously, rather than separately measuring adherence to each medication, to generate one continuous exposure episode. Sometimes, medications within the same disease state but chemically different (e.g., beta-blockers and calcium channel blockers) could be considered interchangeable.

Several types of adherence metrics can be calculated using these data, such as a continuous variable for adherence assessed from the first to last prescription record, a dichotomous variable in which patients are classified as adherent or nonadherent based on a threshold, or examining the time between dispensings. In the most common approach, the proportion of days that patients had an available supply of medication, or the proportion of days covered (PDC), is calculated. The PDC is calculated by dividing the number of days with an available supply of medication by the number of days in the interval being evaluated (an interval-based measure) [48]. Other approaches include calculating the medication possession ratio (MPR). MPR is calculated as the quotient of (1) the total number of days supplied of all dispensings in a given analysis interval for the medication under investigation, and (2) the total number of days in the analysis interval. The primary difference between the PDC and MPR adherence metrics is how overlapping days supplied of the same medication are handled. MPR assesses the total daily medication supply from all dispensings in a given analysis interval whereas PDC assesses the total days where a medication supply "covered" each day in a given analysis interval. The specific approach is typically determined by researcher preference, although may depend to some degree on the structure of the database or the pharmacodynamics of the drug/disease in question. Regardless, the results of the different approaches are typically very similar [23].

In addition, these data can be used to measure persistence (e.g., the time until medication discontinuation) by evaluating whether clinically meaningful treatment gaps or discontinuations are observed in the dispensing data [49]. Potential approaches include evaluating whether a dispensation overlaps with the end of a follow-up period (i.e., 365 days after initiation) or measuring the availability of drug supply at a fixed time after the last medication dispensing (e.g., whether patients have a gap of at least 30, 45, or 60 days with no medication after the supply is presumed to be exhausted). Whichever method is chosen, investigators should conduct sensitivity analyses of the "gap rule" to determine the robustness of the findings.

These adherence measures are limited in additional ways. The vast majority of dispensings for chronic medications in the US are for supplies of 30 days, and increasingly 90 days [47,50]. Measuring adherence in intervals shorter than 180 days can then make it difficult to observe variation in adherence since by definition, the first 30 or 90 days are always considered as full adherence (100%), regardless of actual patient behavior [51]. This problem becomes less pronounced with longer measurement intervals. However, shorter intervals may be more clinically desirable since they might allow nonadherence to be detected and acted upon sooner [52].

Although adherence metrics, using pharmacy dispensing data, often estimate the supply of medication during a given time period, they do not measure or monitor actual pill-taking behavior, either on average or day to day. Consequently, they cannot be used when the timing of missed doses is pivotal. However, the estimation of adherence with pharmacy dispensing data has been shown to be valid for chronic medications where measuring overall exposure between refills is clinically relevant [38]. Pharmacy dispensing measures of adherence have also been shown to be strongly associated with clinical outcomes [53,54]. For example, a time-to-dispensing measure of adherence has been associated with changes in HIV viral load [55]

and changes in blood pressure [44]. Furthermore, the measure has been shown to provide additional information beyond self-reports. In a study of antiretroviral therapy, individuals who self-reported 100% adherence actually varied in their treatment response based on adherence metrics from pharmacy dispensing data. As expected, those with higher adherence, as defined using pharmacy dispensing data, had higher rates of treatment response, despite claims of perfect self-reported adherence in both groups [56].

A limitation of adherence measures derived from pharmacy dispensing data is the estimate of the maximum potential adherence, since these metrics assume all medication supplied has been consumed between dispensations. It is also difficult to disentangle clinically directed medication discontinuation wherein persistence is no longer the behavior being studied, from patient-directed discontinuation against provider recommendation, which is defined as nonpersistence. Furthermore, pharmacy dispensing data may also overestimate adherence measures when dispensing programs automatically dispense a new supply on a prespecified schedule, irrespective of patient request for resupply [57].

A final consideration is how to accurately measure adherence to multiple medications for the same condition (e.g., antihypertensives). One common approach is to measure adherence at the therapeutic class level and "average" adherence across the entire chronic condition for patients exposed to any medication for that condition [27].

Pill Counts

While less commonly used, adherence can also be measured indirectly by pill counts. Pill counts are similar to pharmacy dispensing data in that percent adherence is calculated by dividing the days supply consumed by the number of days observed. Data collected include the dispensing date, quantity dispensed, number of pills per dose, and number of pills left in the bottle, adjusted for doses taken that day and any additional pills left over from the last count.

Like adherence measures estimated using the medication dispensing date and days supplied (e.g., MPR and PDC), adherence measures using pill count data also cannot determine if the medication was actually consumed or the patterns of consumption. However, they do provide direct evidence that the medication was not taken when pills are left over. Pill counts are susceptible to deception since "dumping" pills on the way to the pill count visit is simple and can be done impulsively before a visit. Unannounced pill counts, in person or by telephone, are valid alternatives to mitigate this type of misclassification [55]. During calls, subjects review the contents of each of their pill bottles. Of course, this approach is also susceptible to intentional deception; however, the estimated adherence from pill bottle review was shown to be associated with treatment response [58]. The time for both staff and participants is a potential disadvantage of pill counting and an additional source of error. In addition, missing data can result when patients do not bring in their pill bottles or have them available during telephone calls. Reinforcing the importance of accuracy with staff is vital to ensure validity of this measure.

Medication Diaries

Although the adherence measures described above summarize how much medication was taken over a specified time period, they provide no detail on the timing of missed doses. Depending on drug pharmacokinetics and pharmacodynamics, missing doses may have different consequences depending on whether the doses were missed consecutively or at separate times that are evenly spaced. These data may in fact be vital to classifying adherence, and medication diaries can provide a solution. In this method, participants keep a record of the date and time of each dose of medication and often whether or not it was taken with or without food. These data can be collected either

electronically or handwritten; with newer technologies like smartphones, data collection could be even easier [59]. Diaries may be particularly useful for medications like insulin or inhalers that are difficult to track using other methods [60]. For example, medication diaries are regularly used in pediatric patients [61].

However, medication diaries are susceptible to both overreporting and underreporting of adherence. Social desirability results in patients listing doses even though they were not taken, but the potential is lessened somewhat by the burden of creating a detailed falsified record. In fact, the risk of underreporting may be greater because of the burden of tracking each dose. It is also not easy to employ this method at scale for larger studies. Newer approaches are exploring the use of apps on enabled smartphones to track these more nuanced medication-taking patterns.

Electronic Drug Monitoring Technology

Electronic drug monitors (EDMs) feature the same advantages as medication diaries, but are less susceptible to deception, forgetting, or ignoring the need to write down the dose data. In contrast to the prior approaches, EDMs provide time-stamped data for adherence behaviors to enhance precision of adherence measures. While there are several different hardware options, electronic drug monitors employ electronic date/time stamp technology that is triggered by opening a container (i.e., pill bottle), puncturing a blister pack to obtain a dose, or ingesting a dose. The data are downloaded to a computer or smartphone via hardwired or wireless linkage.

Electronic drug monitor data have been shown to have some correlation with other measures, including pharmacy dispensing and self-report, though EDM measures are more sensitive (i.e., they are more likely to identify poor adherence) than self-reported measures [31,62]. While EDMs are less susceptible to deception than self-report, they could theoretically could be more susceptible than pharmacy dispensing data [63]. However, it is highly unlikely that subjects will open and close the monitor to record doses over long periods of time without actually taking the medication, though this does occasionally happen accidentally [63,64]. EDMs are also less susceptible to underreporting than diaries because they often do not require the subject to do anything other than take the prescribed medication.

Though EDM technology is advancing rapidly, the packaging and cost of EDMs can still be burdensome and difficult to scale [64]. For example, EDMs have been found to be particularly hard for patients with psychiatric conditions to use [64–66]. In addition, they often preclude the use of pillboxes by generally requiring that the medication remain in the package until taken. Consequently, they are susceptible to underestimating adherence (e.g., a one-week supply taken from the container at one time will appear as one dose taken). However, EDMs could be used even when the medication is not kept in the container. In a warfarin study, individuals using pillboxes were given an EDM in an empty pill container and asked to open the empty bottle whenever they took warfarin from the pillbox. The association between adherence and outcome was nearly as strong as those who kept the warfarin in the monitored bottle [20].

Newer approaches are being developed, such as integrating EDM technology with text messaging that reminds patients when they miss doses. In 2015, the US Food and Drug Administration approved the first ingestible sensor technology that measures actual intake time through ingestion of a medication that communicates with an adhesive patch. The device sends a signal to the doctor or research team monitoring adherence [67]. Other research is exploring the utility and accuracy of adherence measures in which patients take date- and time-stamped photographs of themselves or their pills each time they take a dose.

Drug Concentrations

Identification of the presence of a drug in plasma or other tissues provides direct evidence of drug ingestion. However, the use of drug concentrations to measure adherence is limited by variability across patients (i.e., absorption, distribution, metabolism, and clearance – see Chapter 2). The more frequently concentrations are measured, the fuller the picture of adherence behavior that can be obtained but the cost and patient inconvenience may be a limitation. Measurement of drug concentrations in hair using liquid chromatography and confirmed by mass spectrometry can be a useful indicator of long-term medication exposure. For example, antiretroviral drug levels in hair give an average of the exposure to drug over the past weeks to months and predict HIV viral response better than serum drug levels [68].

Unfortunately, many assays are unavailable commercially. Furthermore, the serum drug level is not the relevant measure for many drugs when the site of action is elsewhere (e.g., intracellularly rather than in serum or in hair) [69]. Finally, unless these assays are done quickly, they are not useful clinically.

Another approach to assessing drug concentrations is to use a marker drug that is easily added to a formulation and can be measured more easily than the actual drug of interest. The primary example here is the incorporation of riboflavin into active drugs as a urine metabolite drug marker to assess adherence to medication in clinical trials [70]. Of course, this strategy is only relevant in settings where researchers have control over the formulation and direct access to the patient (e.g., clinical trials).

Measuring Primary Adherence

Each of the approaches described above has focused on later adherence phases (e.g., implementation and persistence). Measuring medication initiation (i.e., primary adherence) has been difficult using some of these methods, particularly because secondary data require a medication to be dispensed for adherence behaviors to be monitored. Some techniques, for example self-report, may allow for easier study of primary adherence. Without linkage to other types of data (e.g., electronic health records that include provider medication orders), it can be difficult to evaluate initiation without knowledge of what was prescribed [5]. Newer approaches are beginning to link these data sources to allow better assessment of the full cascade. On their own, electronic health record data limited to physician orders are not useful at evaluating patient adherence because they do not provide information about medication consumption.

Measuring Adherence to Nonpill Formulations

Measuring adherence to nonpill formulations can be difficult for several reasons, largely because these medications are generally administered with a variable dosing schedule. Injectable medications like insulin may be administered based on a sliding scale, with doses adjusted as needed, so measuring adherence using indirect dispensing data may be imprecise. Recently, other insulin persistence-based measures have been developed to overcome some of these limitations [60,71]. Inhaled medications are also difficult to measure; for those with specific schedules (e.g., tiotropium), dispensing data could be used [72]. Medication diaries and self-report could also theoretically be used but are subject to the same biases as pill formulations. EDMs have been used for metered dose inhalers [73] and ophthalmologic solutions [74]. The monitors increase the size of packaging, but the inhalers and solutions cannot be taken out of the package, unlike pill formulations. Measuring adherence will continue to be a challenge for newer nonpill formulations, including biologics, and in disease states in which both oral and injectable formulations are used interchangeably (e.g., osteoporosis or venous thromboembolism).

Topical treatments pose a particular challenge. For transdermal formulations in patches

(e.g., nicotine, testosterone), adherence based on dispensation data is a viable option because the supply is typically fixed. However, for creams and ointments, because the amount used at each application varies by the size of the lesion being treated or the size of the individual, self-reports and medication diaries may be the only currently viable options [75]. Adherence to intravaginal gels could be monitored by counting the number of empty tubes and used applicators returned at each visit, but this measure is subject to self-report errors due to intentional falsification or mixture of used and unused applicators in the same bag.

Analysis Issues in Adherence

Using Adherence Data in Clinical Trials and Comparative Effectiveness Studies

While clinical trial participants may be more motivated to adhere to treatments than those in clinical practice, nonadherence occurs for all types of self-administered therapies. Missed doses will typically make the active drug less effective and diminish observed differences compared to placebo in intention-to-treat analyses. In order to compensate for this effect, Phase III trials may inflate sample sizes to account for this variability in drug exposure [8]. Clinical trials may also incorporate run-in periods to try to minimize poor adherence (see Chapter 32).

In analyzing trials, the standard approach remains intention to treat. This approach limits the introduction of bias and makes the results more generalizable [76]. However, secondary analyses can be performed on subgroups of adherent patients, but these patients may differ for reasons that may not be easily measurable (i.e., more willing to take therapy, a type of healthy adherer bias) [77,78]. The benefits of randomization would therefore be negated. Moreover, when lifestyle changes are co-interventions along with medication in a trial, the results of secondary analyses will not be true measures of drug efficacy. Of course, medication adherence itself can also be a primary or secondary outcome in randomized trials, particularly for studies of interventions [79–83].

The inclusion of adherence data in analyses of trials is particularly important when a treatment fails. Reasons for failure might include lack of biological effect or lack of adherence. Unless adherence is measured and identified as the cause of failure, the results of the trial will be only partly useful. While regulators will only approve a drug for the studied indication if it is shown to result in improved outcomes, it is important for the drug developer to know if the efficacy of the drug was potentially limited by poor adherence. For example, in one trial, rates of coronary heart disease events were compared in patients randomized to receive either cholestyramine or a placebo [84]. Adherence in the cholestyramine group (defined as taking at least five out of six prescribed packets of cholestyramine per day) was only 50.8% compared to 67.3% adherence in the placebo group due to side effects. Because of the poor adherence, treatment response in the cholestyramine group was attenuated. Thus, because adherence was measured, it was possible to determine that the high rate of intolerable side effects resulted in lower adherence and thus, perhaps, lower treatment effectiveness.

Similarly, observational cohort studies of comparative effectiveness and safety of medications often benefit from measuring and evaluating the relationship between medication adherence and treatment response. First, "as-treated" analyses of safety evaluations often censor follow-up in patients who discontinue therapies for reasons other than toxicity to decrease bias toward the null. Second, marginal structural modeling approaches often include medication adherence as a time-varying exposure. Exploring the relationship between adherence observed between comparators may enrich the conclusions derived from these studies.

Selecting Adherence Intervals

For all adherence measures, a prespecified window for assessing and evaluating adherence must be chosen. The selection of the duration of an adherence interval depends on two important factors: the pharmacokinetics/pharmacodynamics and the granularity of the adherence measurement. For drugs with short half-lives and short onset of action, short intervals are likely to be more clinically relevant than when the drugs have long half-lives and longer onsets of action. For adherence measures that can accurately assess adherence over short periods of time, such as electronic data monitors, shorter intervals can be calculated. By contrast, when measures derived from pharmacy dispensing data are used, adherence analysis intervals must be longer because adherence is based on evaluating patterns between medication dispensing dates in conjunction with the days supplied per dispensing (e.g., 30 days).

The relationship between adherence and outcomes has been well described in antiretroviral therapy and oral contraceptives. Using pharmacy dispensing data, intervals of adherence as long as one year [56] and as short as 30 days [85] have been associated with viral load outcomes with antiretroviral therapy; however, a 90-day measure was found to be more strongly associated with viral load than a 30-day measure. For oral contraceptives, two consecutive days of nonadherence resulted in an unacceptably high rate of treatment failure (i.e., pregnancy) [86].

Unfortunately, for many medications, the most clinically relevant adherence interval may be unknown, and more research is needed to optimize the assessment of adherence. While the choice of interval length depends on the research goals, in general, monitoring adherence over shorter intervals would be desirable, because interventions can be more rapidly implemented. However, shorter intervals are subject to decreased accuracy regarding true adherence behavior. In general, without direct guidance, choices for an adherence interval should be made based on pharmacokinetic and pharmacodynamic data (see Chapter 2).

Statistical Analysis

The simplest approach to summarizing adherence across different methods is the percentage of doses taken (or missed). For electronic monitors, because the timing of each dose is available, percentage of doses taken "on time," standard deviation of time between doses, duration of maximum time gap between doses, and many others can be calculated [4]. For adherence metrics derived from pharmacy dispensing data, the analysis focuses on either the percentage of available medication or the duration of gaps between dispensings [44]. Self-reports focus on the proportion of doses the patients have taken or the time since the last dose was missed [31].

Whichever metric is used, one must choose whether to include adherence as a continuous or dichotomous variable. As previously described, dichotomous thresholds must consider both the likelihood of failure and clinical consequences of treatment failure. Few thresholds have been established based on evidence, yet in research and quality improvement efforts, to dichotomize these adherence variables, patients are often defined as fully adherent if they take at least 80% of prescribed doses. Certainly, many studies in cardiovascular diseases have demonstrated this association; however, there is recognition that different levels of adherence may be required for viral suppression in HIV. In treatment settings with a linear relation between amount of drug taken and therapeutic response, evaluating differences in adherence on a continuous scale would be clinically more meaningful than binary measures. Alternatively, when neither dichotomous nor continuous measures capture the clinically relevant dose–response relationship, assigning ordinal adherence categories (e.g., <70%, 70– < 80%, 80– < 90%, etc.) may be preferable [87].

In addition, evaluating regimens with multiple medications poses analysis challenges [88]. Many classify adherence based on optimal adherence to at least one medication for that disease state (e.g., hypertension) to be fully adherent, although this misclassifies individuals who are nonadherent to some but not all components of the regimen. Fortunately, there is some evidence to suggest that for medications taken simultaneously, adherence to one is highly collinear with adherence to the other [89]. However, differential nonadherence has been documented [90].

Finally, it is difficult to determine whether an individual is poorly adherent or whether the medication is no longer being prescribed when access to medical records is unavailable. This phenomenon poses the greatest challenge for adherence measures derived from pharmacy dispensing data. Further, even when medical records are available and the provider documents the recommendation to discontinue, the exact date of patient medication discontinuation can be difficult to determine.

Time-Varying Nature of Adherence and Trajectory Modeling

Adherence is a nonstatic behavior, and methods are needed to capture changes in adherence over time. This phenomenon has historically been ignored in studies that measure adherence only once or over short intervals. Even when measured longitudinally, adherence data are often averaged. For example, quality measures in the US are based on the proportion of patients with ≥80% adherence (e.g., MPR or PDC) of their prescribed medications over the prior year. However, patients may experience substantial increases and decreases in adherence that are not fully captured by these composite measures.

Consider, for example, one patient who takes a medication perfectly for the first six months but then discontinues for a six-month gap compared with another patient who alternates taking medications perfectly every other week interspersed with week-long gaps over a year. Both patients would have the same calculated adherence (50%) but very different medication use patterns. Composite, cross-sectional measures obfuscate the potential for each patient to experience different health outcomes and require different adherence interventions.

Advanced statistical methods are beginning to take advantage of repeated measurements in adherence data, particularly in dispensing data, to enhance analysis beyond composite measures. One such method applied is group-based trajectory modeling which estimates changes in an outcome that is measured repeatedly over time and identifies individuals with similar longitudinal patterns [91]. In brief, this approach fits a semiparametric (discrete) mixture model and assigns groupings in longitudinal data (e.g., monthly PDC) based on probability distributions for a prespecified number of groups [92]. The probability of belonging to each potential group is modeled as a multinomial logistic regression, and within each group, adherence is modeled as a smooth function of time using up to a fourth order polynomial. The statistical output includes each individual's estimated probabilities of group membership and estimated trajectory curve of adherence over time for each group. For example, a study by Franklin *et al.* of statin initiators identified six distinct patterns of adherence over 15 months, including patients who had (1) near-perfect adherence, (2) poor adherence initially and then improvement, (3) slowly declining adherence, (4) rapid declines in adherence, (5) occasional use, and (6) immediate discontinuation [92].

Researchers ultimately select the best trajectory models based on fit criteria, having sufficient members in each trajectory group (i.e., at least 5% of the overall cohort in each group), narrow confidence intervals and posterior probabilities of membership ≥0.7 (in which each member assigned to that group has at least a 0.7 probability of being in that group) [91].

Trajectory modeling can be accomplished using statistical software with continuous, binary, and count data.

Overall, group-based trajectory models have been shown to summarize adherence patterns better than composite approaches and are strongly associated with clinical outcomes [93]. However, trajectories provide general patterns for adherence behaviors; that is, no one individual follows the exact pattern described by the trajectory of the group to which they are assigned. For example, although a rapid declining trajectory group might be depicted as having adherence decrease starting at month 4, any one individual assigned to that group might begin to be nonadherent at month 3 or month 5. Also, the number of distinct adherence trajectory groups, while guided by the fit criteria above, can also be subjectively based on researchers' interpretations and differ by disease state. Additionally, describing the individual patterns in words can be challenging; labels such as "mid-year discontinuation" and "early nonadherence followed by later return to partial adherence" can be cumbersome. More research is needed to optimize the approach, explore applicability in other adherence data sources, and facilitate communication of findings.

Prediction of Adherence for Interventions

Unfortunately, low rates of adherence have persisted despite extensive efforts to identify and predict patients at risk of poor adherence with the goal of developing interventions to improve adherence. Despite the expansion of databases with rich patient data, prediction of future adherence remains poor. Traditional approaches have focused on clinical and demographic factors at the time of medication initiation, with discriminative ability that is modest at best even with dozens of predictor variables (e.g., c-statistics ranging between 0.6 and 0.7) [52]. Even machine learning, with the capability of measuring complex interactions among predictor variables, has not led to drastically improved prediction, likely because the true factors associated with poor adherence are usually not observable in databases.

One of the more successful approaches has been evaluating patterns of medication filling shortly after initiation. For example, in pharmacy dispensing data, researchers have found that failing to refill in the second and third months after initiation is highly predictive of poor adherence over the following year (i.e., past adherence predicts future adherence) [51]. Predictions of adherence by providers have also been shown to be no better than chance [94], so they should not be used routinely in adherence studies or in practice.

Future Directions

As outlined throughout this chapter, though many methods have been developed to evaluate adherence, many challenges remain. Better methods for detecting and addressing poor adherence as well as the reasons for poor adherence will be welcome developments. Objective measurement of adherence to nonpill formulations in particular is difficult, especially for injectable, liquid, and topical treatments. The optimal adherence metric for most drug–disease dyads remains unknown. This is further complicated by the enormous number of possible combinations of regimens.

Adherence studies are likely to advance in several ways. First, because optimal adherence thresholds may differ across individuals and diseases, researchers are beginning to explore personalized adherence targets. For example, in a machine-learning analysis among patients with diabetes, Lo-Ciganic *et al*. observed that optimal adherence thresholds for an individual's hospitalization risk varied greatly based on their underlying health status [95,96]. Novel approaches using other types of data are likely to emerge as well, including the use of more advanced microelectronic technology, often

linked with communication systems that both identify and report nonadherence, or the enhancement of mobile and smartphone technology for tracking and intervening on adherence. Refinements to currently available electronic monitors will also likely include more convenient packaging that can both help with adherence (e.g., a reminder or organizer system) and provide two-way personalized communication with patients.

Hopefully, with greater recognition of the importance of nonadherence, more research will be conducted over the next several decades to solve some of these problems as well as develop better approaches to improving adherence so that evidence-based medications can be optimally used.

References

1 Iuga AO, McGuire MJ. Adherence and health care costs. *Risk Manag Healthc Policy* 2014; **7**: 35–44.

2 Urquhart J. Defining the margins for errors in patient compliance with prescribed drug regimens. *Pharmacoepidemiol Drug Saf* 2000; **9**(7): 565–8.

3 Steiner JF, Earnest MA. The language of medication-taking. *Ann Intern Med* 2000; **132**(11): 926–30.

4 Vrijens B, de Geest S, Hughes DA, *et al*. A new taxonomy for describing and defining adherence to medications. *Br J Clin Pharmacol* 2012; **73**(5): 691–705.

5 Shrank WH, Choudhry NK, Fischer MA, *et al*. The epidemiology of prescriptions abandoned at the pharmacy. *Ann Intern Med* 2010; **153**(10): 633–40.

6 Cramer JA, Roy A, Burrell A, *et al*. Medication compliance and persistence: terminology and definitions. *Value Health* 2008; **11**(1): 44–7.

7 Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med* 2005; **353**(5): 487–97.

8 Ickovics JR, Meisler AW. Adherence in AIDS clinical trials: a framework for clinical research and clinical care. *J Clin Epidemiol* 1997; **50**(4): 385–91.

9 van Onzenoort HA, Menger FE, Neef C, *et al*. Participation in a clinical trial enhances adherence and persistence to treatment: a retrospective cohort study. *Hypertension* 2011; **58**(4): 573–8.

10 Gellad WF, Grenard JL, Marcum ZA. A systematic review of barriers to medication adherence in the elderly: looking beyond cost and regimen complexity. *Am J Geriatr Pharmacother* 2011; **9**(1): 11–23.

11 Choudhry NK, Denberg TG, Qaseem A. Improving adherence to therapy and clinical outcomes while containing costs: opportunities from the greater use of generic medications: best practice advice from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med* 2016; **164**(1): 41–9.

12 Reynolds NR, Sun J, Nagaraja HN, Gifford AL, Wu AW, Chesney MA. Optimizing measurement of self-reported adherence with the ACTG Adherence Questionnaire: a cross-protocol analysis. *J AIDS* 2007; **46**(4): 402–9.

13 Oyugi JH, Byakika-Tusiime J, Ragland K, *et al*. Treatment interruptions predict resistance in HIV-positive individuals purchasing fixed-dose combination antiretroviral therapy in Kampala, Uganda. *AIDS* 2007; **21**(8): 965–71.

14 Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. *Cochrane Database Syst Rev* 2015; **5**: CD003343.

15 Choudhry NK, Fischer MA, Avorn J, *et al*. The implications of therapeutic complexity on adherence to cardiovascular medications. *Arch Intern Med* 2011; **171**(9): 814–22.

16 Chan DC, Shrank WH, Cutler D, *et al*. Patient, physician, and payment predictors of statin adherence. *Med Care* 2010; **48**(3): 196–202.

17 Claborn KR, Meier E, Miller MB, Leffingwell TR. A systematic review of treatment fatigue among HIV-infected patients prescribed antiretroviral therapy. *Psychol Health Med* 2015; **20**(3): 255–65.

18 Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. *Curr Opin Psychol* 2015; **5**: 31–36.

19 Lauffenburger JC, Shrank WH, Bitton A, *et al*. Association between patient-centered medical homes and adherence to chronic disease medications: a cohort study. *Ann Intern Med* 2017; **166**(2): 81–88.

20 Kimmel SE, Chen Z, Price M, *et al*. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med* 2007; **167**(3): 229–35.

21 Baroletti S, Dell'Orfano H. Medication adherence in cardiovascular disease. *Circulation* 2010; **121**(12): 1455–8.

22 Gross R, Bilker WB, Friedman HM, Strom BL. Effect of adherence to newly initiated antiretroviral therapy on plasma viral load. *AIDS* 2001; **15**(16): 2109–17.

23 Karve S, Cleves MA, Helm M, Hudson TJ, West DS, Martin BC. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. *Curr Med Res Opin* 2009; **25**(9): 2303–10.

24 Little SJ, Holte S, Routy JP, *et al*. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002; **347**(6): 385–94.

25 Furberg CD. The Beta-blocker Heart Attack Trial. *Z Kardiol* 1985. **74**(Suppl 6): 159–63.

26 Papasavvas E, Kostman JR, Mounzer K, *et al*. Randomized, controlled trial of therapy interruption in chronic HIV-1 infection. *PLoS Med* 2004; **1**(3): e64.

27 Choudhry NK, Shrank WH, Levin RL, *et al*. Measuring concurrent adherence to multiple related medications. *Am J Manag Care* 2009; **15**(7): 457–64.

28 Maciejewski ML, Bryson CL, Wang V, Perkins M, Liu CF. Potential bias in medication adherence studies of prevalent users. *Health Serv Res* 2013; **48**(4): 1468–86.

29 Raymond CB, Morgan SG, Katz A, Kozyrskyj AL. A population-based analysis of statin utilization in British Columbia. *Clin Ther* 2007; **29**(9): 2107–19.

30 Krumme AA, Franklin JM, Isaman DL, *et al*. Predicting 1-year statin adherence among prevalent users: a retrospective cohort study. *J Manag Care Spec Pharm* 2017; **23**(4): 494–502.

31 Stirratt MJ, Dunbar-Jacob J, Crane HM, *et al*. Self-report measures of medication adherence behavior: recommendations on optimal use. *Transl Behav Med* 2015; **5**(4): 470–82.

32 Nguyen TM, La Caze A, Cottrell N. What are validated self-report adherence scales really measuring? A systematic review. *Br J Clin Pharmacol* 2014; **77**(3): 427–45.

33 Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 1986; **24**(1): 67–74.

34 Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens* 2008; **10**(5): 348–54.

35 Svarstad BL, Chewning BA, Sleath BL, Claesson C. The Brief Medication Questionnaire: a tool for screening patient adherence and barriers to adherence. *Patient Educ Couns* 1999; **37**(2): 113–24.

36 Wilson IB, Lee Y, Michaud J, Folwer FJ Jr, Rogers WH. Validation of a new three-item self-report measure for medication adherence. *AIDS Behav* 2016; **20**(11): 2700–2708.

37 Amico KR, Fisher WA, Cornman DH, *et al*. Visual analog scale of ART adherence: association with 3-day self-report and adherence barriers. *J AIDS* 2006; **42**(4): 455–9.

38 Hansen RA, Kim MM, Song L, Tu W, Wu J, Murray MD. Comparison of methods to assess medication adherence and classify nonadherence. *Ann Pharmacother* 2009; **43**(3): 413–22.

39 Savitz ST, Stearns SC, Zhou L, *et al*. A comparison of self-reported medication adherence to concordance between Part D claims and medication possession. *Med Care* 2017; **55**(5): 500–5.

40 Stephenson JJ, Shinde MU, Kwong WJ, Fu AC, Tan H, Weintraub WS. Comparison of claims vs patient-reported adherence measures and associated outcomes among patients with nonvalvular atrial fibrillation using oral anticoagulant therapy. *Patient Prefer Adherence* 2018; **12**: 105–17.

41 Lu M, Safren SA, Skolnik PR, *et al*. Optimal recall period and response task for self-reported HIV medication adherence. *AIDS Behav* 2008; **12**(1): 86–94.

42 Gonzalez JS, Schneider HE, Wexler DJ, *et al*. Validity of medication adherence self-reports in adults with type 2 diabetes. *Diabetes Care* 2013; **36**(4): 831–7.

43 Bangsberg DR, Bronstone A, Chesney MA, Hecht FM. Computer-assisted self-interviewing (CASI) to improve provider assessment of adherence in routine clinical practice. *J AIDS* 2002; **31**(Suppl 3): S107–11.

44 Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records. *Description and validation. Med Care* 1988; **26**(8): 814–23.

45 Choudhry NK, Shrank WH. Four-dollar generics – increased accessibility, impaired quality assurance. *N Engl J Med* 2010; **363**(20): 1885–7.

46 Lauffenburger JC, Balasubramanian A, Farley JF, *et al*. Completeness of prescription information in US commercial claims databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(8): 899–906.

47 Thorpe CT, Johnson H, Dopp A, *et al*. Medication oversupply in patients with diabetes. *Res Social Adm Pharm* 2015; **11**(3): 382–400.

48 Benner JS, Glynn RJ, Mogun H, Neumann PJ, Weinstein MC, Avorn J. Long-term persistence in use of statin therapy in elderly patients. *JAMA* 2002; **288**(4): 455–61.

49 Yeaw J, Benner JS, Walt JG, Sian S, Smith DB. Comparing adherence and persistence across 6 chronic medication classes. *J Manag Care Pharm* 2009; **15**(9): 728–40.

50 Lauffenburger JC, Franklin JM, Krumme AA, *et al*. Predicting adherence to chronic disease medications in patients with long–term initial medication fills using indicators of clinical events and health behaviors. *J Manag Care Spec Pharm* 2018; **24**(5): 469–77.

51 Franklin JM, Krumme AA, Shrank WH, Matlin OS, Brennan TA, Choudhry NK. Predicting adherence trajectory using initial patterns of medication filling. *Am J Manag Care* 2015; **21**(9): e537–44.

52 Franklin JM, Shrank W, Lii J, *et al*. Observing versus predicting: initial patterns of filling predicted long-term adherence more accurately than high-dimensional modeling techniques. *Health Serv Res* 2016; **51**(1): 220–39.

53 Bitton A, Choudhry NK, Matlin OS, Swanton K, Shrank WH. The impact of medication adherence on coronary artery disease costs and outcomes: a systematic review. *Am J Med* 2013; **126**(4): 357 e7–357 e27.

54 Choudhry NK, Avorn J, Glynn R, *et al*. Full coverage for preventive medications after myocardial infarction. *N Engl J Med* 2011; **365**(22): 2088–97.

55 Bangsberg DR, Hecht FM, Charlebois ED, *et al*. Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. *AIDS* 2000; **14**(4): 357–66.

56 Low-Beer S, Yip B, O'Shaugnessy MV, Hogg RS, Montaner JS. Adherence to triple therapy and viral load response. *J AIDS* 2000; **23**(4): 360–1.

57 Matlin OS, Kymes SM, Averbukh A, *et al*. Community pharmacy automatic refill program improves adherence to maintenance therapy and reduces wasted medication. *Am J Manag Care* 2015; **21**(11): 785–91.

58 Kalichman SC, Amaral CM, Stearns H, *et al*. Adherence to antiretroviral therapy assessed by unannounced pill counts conducted by

telephone. *J Gen Intern Med* 2007; **22**(7): 1003–6.

59 Mertens A, Brandl C, Miron-Shatz T, *et al*. A mobile application improves therapy – adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation via iPad with paper-based control. *Medicine* 2016; **95**(36): e4446.

60 Stolpe S, Kroes MA, Webb N, Wisniewski T. A systematic review of insulin adherence measures in patients with diabetes. *J Manag Care Spec Pharm* 2016; **22**(11): 1224–46.

61 van Berge Henegouwen MT, van Driel HF, Kasteleijn-Nolst Trenite DG. A patient diary as a tool to improve medicine compliance. *Pharm World Sci* 1999; **21**(1): 21–4.

62 Mehta SJ, Asch D, Troxel A, *et al*. Comparison of pharmacy claims and electronic pill bottles for measurement of medication adherence among myocardial infarction patients. *Med Care* 2019; **57**(2): e9–e14.

63 Lam WY, Fresco P. Medication adherence measures: an overview. *Biomed Res Int* 2015; **2015**: 217047.

64 Checchi KD, Huybrechts KF, Avorn J, Kesselheim AS. Electronic medication packaging devices and medication adherence: a systematic review. *JAMA* 2014; **312**(12): 1237–47.

65 Kozuki Y, Poupore E, Schepp K. Visual feedback therapy to enhance medication adherence in psychosis. *Arch Psychiatr Nurs* 2005; **19**(2): 70–80.

66 Elixhauser A, Eisen SA, Romeis JC, Homan SM. The effects of monitoring and feedback on compliance. *Med Care* 1990; **28**(10): 882–93.

67 Hafezi H, Robertson TL, Moon GD, Au-Yeung KY, Zdeblick MJ, Savage GM. An ingestible sensor for measuring medication adherence. *IEEE Trans Biomed Eng* 2015; **62**(1): 99–109.

68 Bernard L, Vuagnat A, Peytavin G, *et al*. Relationship between levels of indinavir in hair and virologic response to highly active antiretroviral therapy. *Ann Intern Med* 2002; **137**(8): 656–9.

69 Moore JD, Acosta EP, Johnson VA, *et al*. Intracellular nucleoside triphosphate concentrations in HIV-infected patients on dual nucleoside reverse transcriptase inhibitor therapy. *Antivir Ther* 2007; **12**(6): 981–6.

70 Young LM, Haakenson CM, Lee KK, van Eeckhout JP. Riboflavin use as a drug marker in Veterans Administration cooperative studies. *Control Clin Trials* 1984; **5**(4 Suppl): 497–504.

71 Wei W, Pan C, Xie L, Baser O. Real-world insulin treatment persistence among patients with type 2 diabetes. *Endocr Pract* 2014; **20**(1): 52–61.

72 Bryant J, McDonald VM, Boyes A, Sanson-Fisher R, Paul C, Melville J. Improving medication adherence in chronic obstructive pulmonary disease: a systematic review. *Respir Res* 2013; **14**: 109.

73 Julius SM, Sherman JM, Hendeles L. Accuracy of three electronic monitors for metered-dose inhalers. *Chest* 2002; **121**(3): 871–6.

74 Okeke CO, Quigley HA, Jampel HD, *et al*. Adherence with topical glaucoma medication monitored electronically the Travatan Dosing Aid study. *Ophthalmology* 2009; **116**(2): 191–9.

75 Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J, Rask K. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. *J Gen Intern Med* 2006; **21**(12): 1215–21.

76 Albert JM, Demets DL. On a model-based approach to estimating efficacy in clinical trials. *Stat Med* 1994; **13**(22): 2323–35.

77 Dormuth CR, Patrick AR, Shrank WH, *et al*. Statin adherence and risk of accidents: a cautionary tale. *Circulation* 2009; **119**(15): 2051–7.

78 Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive

interventions: a primer for physicians. *J Gen Intern Med* 2011; **26**(5): 546–50.

79 Viswanathan M, Golin CE, Jones CD, *et al*. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. *Ann Intern Med* 2012; **157**(11): 785–95.

80 Choudhry NK, Isaac T, Lauffenburger J, *et al*. Effect of a remotely delivered tailored multicomponent approach to enhance medication taking for patients with hyperlipidemia, hypertension, and diabetes: the STIC2IT Cluster Randomized Clinical Trial. *JAMA Intern Med* 2018; **178**(9): 1182–9.

81 Kimmel SE, Troxel AB, Loewenstein G, *et al*. Randomized trial of lottery-based incentives to improve warfarin adherence. *Am Heart J* 2012; **164**(2): 268–74.

82 Choudhry NK, Krumme AA, Ercole PM, *et al*. Effect of reminder devices on medication adherence: the REMIND Randomized Clinical Trial. *JAMA Intern Med* 2017; **177**(5): 624–31.

83 Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980; **303**(18): 1038–41.

84 Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. *JAMA* 1984; **251**(3): 351–64.

85 Acri TL, Grossberg RM, Gross R. How long is the right interval for assessing antiretroviral pharmacy refill adherence? *J AIDS* 2010; **54**(5): e16–e18.

86 Smith SK, Kirkman RJ, Arce BB, McNeilly AS, Loudon NB, Baird DT. The effect of deliberate omission of Trinordiol or Microgynon on the hypothalamo-pituitary-ovarian axis. *Contraception* 1986; **34**(5): 513–22.

87 Gross R, Bellamy SL, Chapman J, *et al*. Managed problem solving for antiretroviral therapy adherence: a randomized trial. *JAMA Intern Med* 2013; **173**(4): 300–6.

88 Lauffenburger JC, Landon JE, Fischer MA. Effect of combination therapy on adherence among US patients initiating therapy for hypertension: a cohort study. *J Gen Intern Med* 2017; **32**(6): 619–25.

89 Feldman HI, Hackett M, Bilker W, Strom BL. Potential utility of electronic drug compliance monitoring in measures of adverse outcomes associated with immunosuppressive agents. *Pharmacoepidemiol Drug Saf* 1999; **8**(1): 1–14.

90 Gardner EM, Burman WJ, Maravi ME, Davidson AJ. Durability of adherence to antiretroviral therapy on initial and subsequent regimens. *AIDS Patient Care STDS* 2006; **20**(9): 628–36.

91 Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol* 2010; **6**: 109–38.

92 Franklin JM, Shrank WH, Pakes J, *et al*. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. *Med Care* 2013; **51**(9): 789–96.

93 Franklin JM, Krumme AA, Tong AY, *et al*. Association between trajectories of statin adherence and subsequent cardiovascular events. *Pharmacoepidemiol Drug Saf* 2015; **24**(10): 1105–13.

94 Gross R, Bilker WB, Friedman HM, Coyne JC, Strom BL. Provider inaccuracy in assessing adherence and outcomes with newly initiated antiretroviral therapy. *AIDS* 2002; **16**(13): 1835–7.

95 Lo-Ciganic WH, Donohue JM, Thorpe JM, *et al*. Using machine learning to examine medication adherence thresholds and risk of hospitalization. *Med Care* 2015; **53**(8): 720–8.

96 Lo-Ciganic WH, Donohue JM, Jones BL, *et al*. Trajectories of diabetes medication adherence and hospitalization risk: a retrospective cohort study in a large state Medicaid program. *J Gen Intern Med* 2016; **31**(9): 1052–60.

39

Risk Evaluation and Communication

Susan J. Blalock[1], Rebecca Dickinson[2], and Peter Knapp[3]

[1] *Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
[2] *School of Health and Community Studies, Leeds Beckett University, Leeds, UK*
[3] *Department of Health Sciences and the Hull York Medical School, University of York, York, UK*

All medications have risks. Although many different definitions exist, risk is usually defined as a potential harmful outcome that can occur with a known or unknown probability [1]. Some medication-related risks are more serious than others and some are well understood whereas others are clouded by uncertainty. The responsibility of ensuring that medications are used as safely as possible is shared by the pharmaceutical companies that develop, investigate, manufacture, and market medications; the governmental agencies charged with regulating these processes; the healthcare providers who prescribe or dispense prescription medications and make recommendations concerning the use of over-the-counter products; the governmental agencies that license and regulate healthcare providers and healthcare facilities; and the patients who ultimately must decide whether or not to use a medication and, in most cases, have control over how they use the medication.

Since passage of the Kefauver–Harris Amendments to the Food, Drug, and Cosmetic Act in 1962, market approval of a new drug in the United States has required that the Food and Drug Administration (FDA) determine that the medication is safe and effective (see Chapter 1) [2]. Similar criteria are used by regulatory agencies in other countries as well [3]. For example, in Europe, the European Medicines Agency (EMA) is responsible for the scientific evaluation of medicines for the 28 member states of the European Union. Similar to the FDA, the EMA is required to evaluate whether medications are acceptably safe and effective prior to drugs being permitted a marketing authorization or product license [4]. Other chapters in this book provide information concerning how these determinations are made (see Chapters 1 and 8). Here, we simply reiterate that even medications that are judged as meeting safety standards have risks.

A drug is considered "safe" if the risks associated with it are deemed to be acceptable [5]. In some cases, medications with substantial and serious risks are judged as meeting safety standards because the benefits of the medication outweigh the risks. This is most often the case for medications used to treat debilitating or life-threatening illnesses where few other effective treatment options are available. It is also important to recognize that the safety of a medication is not solely an inherent property of the medicine, but also the circumstances in which the medication is used (e.g., expertise of

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

prescribers, procedures used to monitor potential adverse effects, presence of interacting medications). Thus, many medication risks may be minimized through the implementation of appropriate risk management strategies.

To minimize medication risks following market approval, all parties involved in the medication use process must have access to up-to-date information concerning potential risks, including measures that can be used to prevent or control these risks. Moreover, this information must be provided in a timely manner and in a way that is understood by the target audience and that facilitates informed decision making. In this chapter, we discuss some of the clinical and methodologic challenges that must be addressed to meet these goals. We also discuss approaches that are currently used to enhance the dissemination and usability of information concerning medication risks. We conclude by suggesting directions for future research in this area.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Five major clinical issues involving medication risk communication need to be addressed by pharmacoepidemiologic research.

First, one must determine the information that patients need about medication risks to be able to participate in shared decision making and use medications safely. Most medications have many risks (e.g., stomach upset, liver toxicity, cancer, potential for allergic reactions) and each risk has many dimensions that can affect judgments of acceptability. These dimensions include probability, severity, controllability, reversibility, and time of onset (e.g., whether potential harm usually occurs soon after initiation of therapy or may not arise for many years) [6,7]. In addition, although uncertainty is an inherent characteristic of any risk, the risks associated with some medications are more uncertain than others. For example, the risks associated with medications that have been used for many years in a large number of patients may be fairly well understood [8]. Conversely, we often have limited understanding of the risks associated with recently marketed medications, particularly those that are first-in-class, and previously unrecognized risks may continue to emerge for several years after a medication is first marketed. Given that it is probably neither feasible nor desirable to provide comprehensive patient education concerning all medication risks, there is a need to determine how different types of information should be prioritized.

Second, one must determine what information patients need about potential medication *benefits* to be able to make informed decisions regarding the need for therapy and the selection of a specific treatment when therapeutic alternatives are available. Serious risks associated with a particular medication may be acceptable if the medication offers substantial benefits, especially if no acceptable therapeutic alternatives are available [7]. However, the same risk may be unacceptable for a less effective medication that does not provide unique advantages over other treatment options (see Chapter 35).

Third, there is a need to identify the most appropriate targets (e.g., healthcare providers, patients with a specific health problem or taking specific medications, consumers in general) for different types of communications and the most feasible and cost-effective way to communicate the information needed by different target audiences. Although patients can obtain information about medication risks and benefits from a wide variety of sources, most express a preference for obtaining this type of information from their healthcare provider [9,10]. Yet healthcare providers often struggle to remain abreast of recent research findings given the sheer volume

of emerging information as well as conflicting findings from different studies [11]. In addition, physicians may lack the skills in evidence-based medicine needed to critically evaluate the literature [12,13].

Fourth, one must determine how information should be tailored to individuals' needs, preferences, abilities (e.g., health literacy and numeracy), risk status (i.e., presence of factors that affect the probability/severity of medication side effects; presence of factors that affect benefits that might be gained by using the medication and risks associated with deciding to forego therapy), and current status in the medication use process (e.g., deciding whether to initiate therapy with a new medication; self-managing a stable, chronic medication regimen). Tailoring is also needed when working with special populations (e.g., children, individuals with cognitive impairments).

Finally, there is a need to address ethical issues that arise when communicating information about medication risks and benefits. Potentially, educating patients about medication risks and the uncertainty associated with experiencing medication benefits may increase patient reluctance to use an effective prescribed medication [14].

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

In addition to the clinical issues described above, five major methodologic issues need to be addressed by pharmacoepidemiologic research. First, effective risk communication requires the availability of high-quality information concerning the risks and benefits of different therapeutic options, including the option of foregoing treatment. This need goes beyond simply knowing that a certain risk/benefit is possible. Information is needed concerning all the risk dimensions discussed in the previous section (e.g., probability, severity, controllability, reversibility, time of onset). Moreover, the information included in risk communications must be relevant to the target audience. Thus, the generalizability (and limits to generalizability) of pharmacoepidemiologic studies must be well understood.

Second, there is a need to determine the best format for providing risk/benefit information. Most risk communications include probabilistic information, which even healthcare providers can find difficult to interpret [15–17]. A wide variety of formats can be used to convey probabilistic information: qualitative descriptors (e.g., common, rare), numbers (e.g., absolute risk, relative risk, odds ratios), and graphics (e.g., bar charts, pictographs). Many studies have demonstrated that the format used to express probabilistic information can have a substantial impact on judgment and decision making [18,19]. Experts recommend against providing risk information only in relative terms, isolated from baseline rates and other information that would contextualize the risk [20,21]. Experts also recommend that verbal descriptors should either be avoided or defined explicitly in numerical terms as part of the risk communication [22,23]. However, many questions remain concerning the optimal way to convey this type of information.

Third, there is a need to better understand the factors that influence individual differences in how people perceive and respond to risk. Risk evaluation is not simply a cognitive exercise where estimates of probability and severity are entered into a mathematical formula to derive an estimate of acceptability that is invariant across individuals. People respond affectively to risk information [24–27] and different people may respond differently to the same information based on their past experiences and tolerance of uncertainty.

Fourth, there is a need to develop communication strategies that address the tendency for

patients and providers to overestimate the probability and magnitude of medication benefits. Hoffman and Del Mar identified a "medical optimism" among patients who express overly optimistic expectations about interventions, while simultaneously underestimating the chance of harm associated with treatments [28]. In comparison, it might be expected that clinicians have more accurate expectations of the benefits and risks associated with treatments. However, it has been demonstrated that, similarly to patients, clinicians rarely hold accurate expectations of treatment benefits and harms and also tend to overestimate benefit and underestimate risk [29]. Potential reasons for this may include a tendency to make decisions based on an understanding of how a treatment works as opposed to how effective it is [30] or may result from deficits in training. Hoffman and Del Mar also propose the existence of therapeutic illusion, "an unjustified enthusiasm for treatment on the part of both doctors and patients," as a factor that may influence perceptions [29]. Clinicians' understanding of risk and benefits is essential to ensure that patients receive accurate information to make unbiased and informed decisions about their treatments, so this is clearly an issue that requires resolution.

Fifth, there is a need to determine the most appropriate methods to use when evaluating communication effectiveness. In most cases, the ultimate objective of risk communications is to improve health outcomes by reducing the incidence of adverse events. However, it is helpful to consider the causal mechanisms through which desired effects on health outcomes might be achieved. As shown in Figure 39.1, the most proximal effects of risk communications are likely to be increased knowledge and, in many cases, emotional arousal. Next, the information communicated may be incorporated into decision-making processes. At this point, it is important to consider whether the purpose of the risk communication is to inform or persuade. If the purpose of the communication is purely informational, it would not be appropriate to evaluate message effectiveness in terms of the specific decision made. However, many risk communications include components that advocate specific actions (e.g., initiating precautionary behaviors to reduce a specific risk) and, therefore, have a persuasive intent. In these cases, the message would probably not be considered effective unless the desired behavior changes were realized. Evaluators must also give careful consideration to the time required for different types of effects to become evident [31]. For example, one would expect knowledge change to be evident immediately following message exposure. However, effects on health outcomes may require substantial time to become evident.

Two additional issues related to the evaluation of communication effectiveness deserve special attention. First, when evaluating the effect of risk communications on knowledge, investigators must determine what knowledge is needed for patients to make informed decisions and use medications safely. For example, is it important for patients to know that the probability of experiencing a particular medication side effect is 1%? Or is it sufficient to know that the side effect is possible, but unlikely? Reyna has argued that informed decision making does not require recall of precise verbatim information (e.g., exact probabilities), but does require understanding and recall of the essence of the information communicated [32,33]. Second, when evaluating communication effectiveness, it is important to consider unintended as well as intended effects. Risk communications concerning one medication have the potential to raise concern about unrelated medications and result in patients discontinuing efficacious medications that pose minimal risks. Unintended consequences might best be evaluated by assessing changes in health-related quality of life and changes in the use of medications other than those that are targeted by the risk communication.

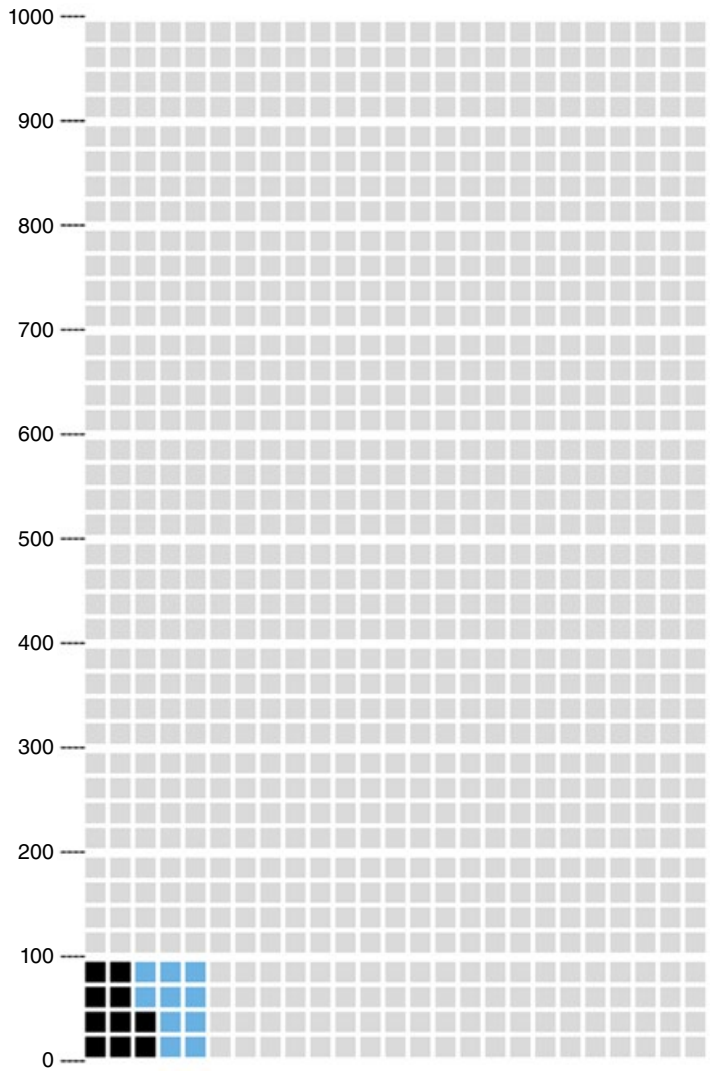


Figure 39.1 Sample pictograph.

Currently Available Solutions

Health Professional and Consumer Medication Information

Many countries have implemented regulatory measures in response to the challenges of communication about medication risks and benefits. For example, in order to promote the principles of transparency and develop methods of improving risk communication, the EMA requires the production of several documents which all play a role in risk/benefit communication. These include the European Public Assessment Report (EPAR), the Risk Management Plan (RMP), and the Patient Information Leaflet (PIL).

The EPAR is a lengthy public document which details the scientific assessment of a pharmaceutical product. These regulatory documents

are written for professionals, but pharmaceutical companies must also provide a user-friendly lay summary. The aim of the lay summary is to communicate how the decision to license the medicine was made. It is essential to ensure that such documents can be understood by potential users, and user testing is one potential method for ensuring that readability is optimized [34].

To adhere to good pharmacovigilance guidelines set out by the EMA, medicinal products are authorized "on the basis that in the specified indication(s), at the time of authorisation, the risk-benefit balance is judged to be positive for the target population" [35]. The EMA requires that pharmaceutical companies publish RMPs which must include information on the medicine's safety profile, how any risks will be prevented or minimized in patients, any plans for studies or other activities to gain more knowledge about the safety and efficacy of the medicine, and to assess the effectiveness of risk minimization measures. There is a requirement that RMPs are updated and modified as needed throughout the lifetime of the medicine, as additional information becomes available.

Patients in the UK and throughout the countries of the European Union must, by law, receive written information with their medicine which includes some communication on the risks and benefits associated with taking it. In 1992, a Directive from the European Commission on the labeling of medicinal products for human use mandated that all medicines are accompanied by a regulated patient information leaflet (package insert) inside the medicine box [36]. The Directive aimed to provide a standardized format for patients in order to provide consumer protection and ensure access to full and comprehensible information about medicines. For some patients, this might be the only written information they receive about their medicines.

European PILs follow a standardized format and include information about the following: potential side effects and their estimated frequency; what the medicine does and what it is for; dos and don'ts; and how to take the medicine [37]. Effective and balanced risk communication occurs when there is presentation of information about both the risk of harm and likelihood of benefit associated with taking a treatment. European PILs currently contain both textual and numerical descriptions of the estimated frequency of harms associated with taking the drug, as well as an indication of severity, aiming to inform patients about potential adverse effects in order to encourage help seeking, but also to support patients in making informed decisions about treatments.

Initially, guidance required European pharmaceutical license holders to present the risk of harm using a combination of qualitative and quantitative descriptions for five bands of risk frequency, ranging from "very common" (>10% frequency) to "very rare" (<0.01%). While the inclusion of the risk frequencies meets patients' identified need for information about the side effects of their treatments [38], the use of percentages to communicate numerical information about risk of harm can lead to overestimations of risk [39], which have been noted both in patients and in doctors [40]. Consequently, updated guidance recommends the use of frequency bands using a natural frequency numerical format (e.g., "less than 1 in 100"). UK PILs now communicate risk using the following regulatory standard [41].

- Very common – occurs more frequently than 1 in 10 administrations of a drug.
- Common – occurs in between 1 in 10 and 1 in 100 administrations of a drug.
- Uncommon – occurs in between 1 in 100 and 1 in 1000 administrations of a drug.
- Rare – occurs in between 1 in 1000 and 1 in 10 000 administrations of a drug.
- Very rare – occurs in less than 1 in 10 000 administrations of a drug.
- Frequency not known.

Use of the five verbal terms (very rare; rare; uncommon; common; very common) alone has

been shown in a number of studies to produce risk estimates that are inconsistent with the assigned frequencies of incidence. For example, "common" was assigned to incident rates between 1% and 10% but was found to result in mean risk estimates of 45.3% by members of the public asked about a hypothetical antibiotic [40]. Furthermore, the term "rare," which had been assigned to rates of 0.01% to 0.1%, produced average risk estimates of 8%. These findings were replicated in a study with a similar study design but which asked people using statin (cholesterol-lowering) medicines to respond to information about real side effects (and their incidences). The "common" side effect of constipation was estimated to occur in 34.2% statin users, while the "rare" side effect of pancreatitis was estimated to occur in 18% users [42]. If these significant overestimates of risk frequency were translated into behaviors (such as deciding not to take the medicine), then their use would be problematic.

One notable point from these and other studies of risk estimation is the high levels of variation in estimates found in study samples. In the study by Berry *et al.*, the standard deviations around the estimates of "common" and "rare" were 22.5% and 7.5%, respectively, a pattern replicated elsewhere [42]. Risk estimates among people are highly variable, which may in part result from relatively stable differences between them in their perceptions of risk susceptibility or their numeracy skills [43]. However, verbal terms appear to add another layer of variation and it seems to be much more difficult to achieve consensus in their meaning than it is for numerical risks descriptors.

Verbal terms do have some strengths, as they may be seen as less intimidating by some patients and as closer to the everyday language of risk; in conversation, people will tend to use words to give a gist or estimate of the degree of risk associated with an uncertain event. However, when used in the context of medicine side effects, words seem to have framing effects, tending to produce inflated estimates of risk that might lead patients to make inaccurate judgments [44]. PILs, as produced currently according to European law, are perceived as containing information about medicines as risky, side effect-inducing products [45]. Therefore, it is important that the terms used, both in writing and in conversation with patients, are proportionate and not prone to misinterpretation.

When space allows, using graphical representations of risk (e.g., bar charts, pictographs) can be helpful, although they are not a panacea and evaluation is essential. However they do seem to have positive effects on accuracy of risk estimates, as well as other benefits, and appear particularly useful for information users who are less skilled or adept at risk interpretation [46]. A sample pictograph is shown in Figure 39.1. In this graph, each square corresponds to one person in the at-risk population. The black squares depict the risk of experiencing liver toxicity within five years of initiation of therapy with Medication B. Thus, this figure suggests that 10 out of every 1000 patients taking Medication B will experience liver toxicity within five years of therapy initiation. The dark gray squares depict the increase in risk associated with Medication A. That is, out of every 1000 patients treated with Medication A, 10 extra cases of liver toxicity would be expected to develop among individuals taking Medication A as opposed to Medication B. Pictographs provide a relatively simple way to convey information concerning both relative and absolute risk. However, they may not be useful when the risk(s) of interest occur very infrequently (e.g., 1 case/10 000 patients treated).

One criticism that patients express about European regulated PILs is that, while information about side effects is valued, when included in written information it can provide an overly negative impression of the medicine. Consequently, in order to support patients to make balanced and informed decisions about treatments, there is a need to also provide information about the

likelihood of benefit from the treatment alongside the likelihood of harm. Currently, there is no consensus on the best format for the presentation of this type of positive information although the EMA notes that benefit information should be compatible with the Summary of Product Characteristics, a document required for the licensing of a drug, and should not be promotional.

The inclusion of information about the potential benefits of medicines in written medicines information is not without its challenges. As noted previously, there is a tendency for patients to overestimate the benefits of their treatments [28] and this can translate to "concern and upset" for patients when they are provided with numerical information that contextualizes the likelihood of benefit they receive from their treatments. A series of qualitative studies exploring patient perceptions and opinions on the provision of textual and numerical benefit information in patient information leaflets have all presented similar findings.

Hamrosi and colleagues recruited focus groups of medicine users in the UK and Australia and provided information about the medicine clopidogrel in two different formats, which had been revised specifically for the study [47]. The revised leaflets contained additional benefit information about clopidogrel in either a text-only format or a numerical format based on the number needed to treat (NNT) to prevent a heart attack or stroke. The NNT statement was written as follows, based on the best available clinical trial data:

> "If 100 people took this medicine for 2 years:
> - 3 of them would be saved from having a heart attack
> - 1 of them would be saved from having a stroke"

A key finding from this study was that while the inclusion of benefit information was valued and seen as a positive addition to the leaflet, the numerical benefit information provoked strong feelings of shock and surprise at the perceived low chance of benefiting from treatment. Some participants did not understand the numerical information, while others struggled to comprehend the magnitude and made an assessment of potential benefit based on a crude interpretation of the data.

Similar research has been undertaken exploring patient opinion and perception of different presentations of benefit information for medicines with different magnitudes of risk. In a focus group, study participants were presented with benefit statements for two different drugs (i.e., sumatriptan and simvastatin) using one of three different numerical formats (percentages, natural frequencies, and NNT) [48]. An example of the different magnitude of benefit seen in the two drugs is presented as follows using the NNT format.

- If 4 people like you take sumatriptan, 1 of them will have a less severe migraine headache after 2 hours.
- If 20 people like you take simvastatin over the next 5 years, 1 of them will be stopped from having a heart attack or stroke.

Participants reported similar levels of shock and surprise at the perceived poor benefits from the two medicines. Other key findings included the following:

- Textual format information was preferred, but did not provide enough information to help contextualize the magnitude of benefit.
- The NNT format was frequently misunderstood.
- The natural frequency format was challenging to understand but when participants invested time to understand, they reported that it helped understanding.
- Numerical information was perceived as worrying, but it was valued.
- Some participants thought that if information on the chance of benefit is available, it should

not be withheld from patients, who may want it to help with their decisionmaking.

One concern noted was that numerical benefit information has the potential to influence patient behavior and could lead to the rejection of a beneficial treatment, perhaps based upon the affect heuristic, a mental shortcut that results in decisions being made based on an emotional response to information rather than reason [49].

There is evidence that "patients" shown numerical benefit information are more likely to choose not to take a treatment, rather than when they are shown nonnumerical information [50]. The inclusion of numerical information on the chance of benefit can influence decision making resulting in a tendency to reject a treatment. This study also explored the impact of the provision of numerical information about side effects and found the opposite – when individuals are provided with nonnumerical side effect information, they are less likely to take a medication than those provided with numerical information about the likelihood of harm. A key finding from this study is that presenting side effect and benefit information in nonnumeric format appears to influence decision making in opposite directions. Although numeric information for both benefits and side effects may enhance decision making, providing numeric benefit information may decrease individuals' willingness to take the medicine, creating both an ethical dilemma for prescribers and providers and a public health concern for policy makers when the chance of benefit from a medicine makes its use attractive at a population level, but which may not be persuasive for individual patients.

In the United States, all prescription medications are required to have an FDA-approved package insert, targeted primarily for prescribers, that comprises the official product label. However, with a limited number of exceptions, there are no regulations that require patients to receive written information about medication risks/benefits when they obtain prescription medications. This is despite FDA recognition that:

> … people are able to make better decisions about their healthcare and better use of the prescription medications available to them when they are well informed about the medications they take. Access to useful written information about prescription medications is important to ensuring appropriate use of these products [51].

Since the late 1960s, the FDA has required that patients receive a patient package insert when they obtain prescriptions for oral contraceptives and estrogens [51]. In 1979, the FDA proposed regulations requiring that manufacturers develop and distribute written patient information, to be approved by the FDA, for all prescription medications. However, the regulations were revoked prior to implementation, based in part on assertions made by pharmaceutical manufacturers and other private sector stakeholders that the goals of the regulations could be accomplished without governmental regulation. Unfortunately, although the availability of consumer medication information has increased over the past 40 years, the quality of the materials distributed is variable and often poor [52]. For example, a study reported in 2007 found that although most pharmacies in the US distribute written materials with prescription medications, many of the materials failed to include information such as contraindications and precautions needed for safe medication use [53]. Notably, there was considerable variability in the consumer medication information distributed by pharmacies in the three countries examined: the US, Australia, and the UK. The materials distributed in the US were evaluated the least favorably.

In an attempt to assist private sector developers, the FDA issued a Guidance document in 2006 entitled "Guidance on Useful Written Consumer Medication Information (CMI)" [51].

Table 39.1 FDA Action Plan criteria for defining useful consumer medication information.

Criterion
Drug names, indications for use, and how to monitor for improvement
Contraindications and what to do if they apply
Specific directions about how to use and store the medicine, and overdose information
Specific precautions and warnings about the medicine
Symptoms of serious or frequent possible adverse reactions and what to do
Certain general information, including encouraging patients to communicate with healthcare professionals, and disclaimer statements
Information that is scientifically accurate, unbiased in tone and content, and up to date
Information in an understandable and legible format that is readily comprehensible to consumers

Source: FDA [55].

Although this document does not establish regulations or legal requirements, it does provide recommendations for the content and format of CMI. As shown in Table 39.1, the Guidance document identifies eight criteria that can be used to assess the usefulness of CMI. It recommends that CMI not include all possible side effects but rather focus on those that are the most serious and most common. The Guidance document does not include any recommendations concerning how to communicate information concerning the likelihood of experiencing the side effects included. However, the examples provided in the Guidance suggest that no numerical information is needed. For example, in a section labeled "Possible side effects," the sample CMIs included in the Guidance state: "The most common side effects are mild upset stomach, diarrhea, and rash. Call your health care provider if these side effects bother you or do not go away" [51]. Finally, although the Guidance document highlights the need to write CMI using plain language principles, it does not recommend user testing to assess consumer comprehension.

A study conducted in 2008 assessed the extent to which CMI distributed by retail pharmacies in the United States met these criteria [54]. This study found that although 94% of the pharmacies visited by secret shoppers provided CMI with prescriptions for lisinopril and metformin, the materials met only about 60% of the eight criteria specified in the FDA Guidance document. Moreover, less than 50% of the materials were judged as meeting the criteria for comprehensibility/legibility, leading the investigators to conclude that "Private sector initiatives to provide useful CMI have failed."

The FDA Amendments Act (FDAAA-PL 110-85) of 2007 gave the FDA authority to require that pharmaceutical manufacturers submit a Risk Evaluation and Mitigation Strategy (REMS) to the FDA when deemed necessary to ensure that the benefits of a drug or biologic product outweigh its risks [55]. The FDA may require a manufacturer to submit a REMS as part of the initial drug approval process or in response to a new safety concern identified via sources such as adverse event monitoring systems, peer-reviewed biomedical literature, clinical trials, and the FDA's Sentinel Initiative [56]. As of February 2018, 73 products have an approved REMS.

All REMS must include at least one safety-related goal that identifies the specific health outcome that the REMS is designed to accomplish [55]. For example, the goal of the Prolia® (denosumab) REMS is to:

> … mitigate the risks of hypocalcemia, osteonecrosis of the jaw, atypical femoral frac-

tures, serious infections, and dermatologic reactions by:

1) informing healthcare providers and patients about the risks of (1) hypocalcemia, (2) osteonecrosis of the jaw, (3) atypical femoral fractures, (4) serious infections, and (5) dermatologic reactions associated with PROLIA®.
2) informing healthcare providers they should counsel patients about the risks associated with PROLIA® [57,58].

Risk Evaluation and Mitigation Strategies may include three major components [55]. First, the manufacturer may be required to develop a Medication Guide or a patient package insert which must be given to patients when a prescription is filled. Medication Guides must be approved by the FDA and become part of the official drug label. Approximately half of the currently approved REMS include a Medication Guide. However, the FDA can require a Medication Guide for drugs and biologic products that do not have a REMS if they determine that "certain information is necessary to prevent serious adverse effects, patient decision making should be informed by information about a known serious side effect with a product, or patient adherence to directions for the use of a product are essential to its effectiveness" [59]

Currently, approximately 600 products have Medication Guides. A 2012 study [60] examined all the Medication Guides that were available in April 2010 to determine the extent to which they met criteria of suitability for use among individuals with limited literacy skills [61]. Of the 185 Medication Guides assessed, only one was deemed suitable for individuals with low literacy skills. In a separate substudy, the investigators asked study participants to review three Medication Guides (taken one at a time) and their comprehension of the information contained in the Guides was assessed using "open book" methods. Participants answered an average of only 52.7% (SD 22.6) of the comprehension questions correctly, with lower scores observed among those with low and marginal literacy. Another study analyzed the results of 66 unique Medication Guide assessments submitted to the FDA between September 2008 and June 2012 [62]. On average, participants correctly answered 63.5% of questions concerning the primary drug risk(s). Only 20 Medication Guide assessments (30.3%) reported knowledge scores of 80% or higher. In general, higher knowledge scores were reported for Medication Guides that were part of a REMS that also included either a Communication Plan or Elements to Assure Safe Use, as described below. Other studies have also demonstrated that many patients have difficulty understanding the information contained in Medication Guides, including the critical safety information included in these documents [63,64]. Thus, although the number of Medication Guides available has expanded dramatically over the past decade, work to improve the usability of these Guides is urgently needed.

Second, a REMS may be required to include a Communication Plan targeted at healthcare providers [66]. For example, the Communication Plan for Prolia targets healthcare providers likely to prescribe this medication [67] and includes: (1) a letter to healthcare providers; (2) a letter to professional societies; (3) a patient counseling toolkit that includes a patient counseling chart for healthcare providers, a patient brochure, and a Medication Guide; (4) a journal information piece that was published quarterly for 12 months in three targeted journals; (5) a plan for the dissemination of REMS information at scientific meetings; and (6) a REMS Website that provides access to all the materials included in the REMS [68]. In line with FDA requirements, the landing page of the REMS website includes a statement encouraging patients and healthcare providers to report suspected adverse reactions and provides a link to the MedWatch reporting system (see Chapter 10), as well as toll-free telephone numbers to both the FDA and the manufacturer.

The final REMS component involves elements to assure safe use (ETASU). These elements

may include specific training, experience, or certifications for healthcare providers who prescribe or dispense the drug; restricting the types of healthcare settings in which the drug can be dispensed; special requirements for patient monitoring; documentation of required safety measures (e.g., laboratory testing); and patient enrollment in a drug registry.

Manufacturers are required to evaluate the effectiveness of their REMS 18 months, three years, and seven years after the REMS is approved. The results of these evaluations must be reported to the FDA so that it can determine whether modifications to the REMS are needed. Morris has provided guidelines for the assessment of REMS programs [69]. In 2011, the FDA created a REMS Integration Initiative to better understand the overall effectiveness of the REMS requirements and identify ways in which current regulations might be improved, particularly in ways that would reduce burdens associated with the regulations while not compromising the effectiveness of the program [70].

Patient–Provider Communication

Although healthcare providers have a professional obligation to counsel patients about medication risks, they may be reluctant to discuss potential risks with patients due to concern that it may decrease patient adherence to the prescribed medication regimen [71]. However, research suggests that the opposite is true. Patient–provider communication concerning potential medication risks and incorporation of patient preferences into the decision-making process may increase adherence and decrease the likelihood of premature discontinuation of therapy [72–74]. Unfortunately, research suggests that this type of communication during patient office visits is not the norm. For example, Sibley and colleagues found that medication concerns (e.g., expected side effects) were discussed in only 2.7% of visits involving diabetes patients and a nurse prescriber [75]. In another study, Richard and Lussier found that potential adverse reactions were discussed in fewer than 17% of physician office visits in which a new medication was prescribed [76].

In a study that analyzed audiotaped visits of patients with rheumatoid arthritis and their rheumatologist, Blalock and colleagues found that, when medication risks were discussed in relation to a medication that was being proposed for addition to the patients' regimen, the types of information most frequently provided were the importance of monitoring to detect potential problems early (30%), probability of side effect occurrence (29.8%), steps to take to minimize risk (25.5%), and severity (21.8%) [77]. When discussing risks associated with medications the patient was currently taking, only the importance of monitoring and steps to take to minimize risk were discussed in over 20% of the conversations. These findings highlight that patient information needs vary depending on their stage in the medication use process (e.g., deciding whether to initiate a new medication, managing a current medication regimen). This study also found that patients often were not able to extract meaningful gist from the information communicated by the rheumatologists [78]. For example, in 14% of the visits, patient coders indicated that it was not clear if the rheumatologist thought the medication was needed and, in 29% of the visits, the coders indicated that it was not clear if the rheumatologist was concerned about the safety of the medication. These findings highlight the need to focus not only on the types of risk/benefit information communicated, but also on the clarity of the communication.

Pharmacists also have a professional obligation to counsel patients about medication risks. Internationally, the rates of verbal counseling provided by pharmacists in community pharmacy settings tend to be low, but vary widely depending upon the research methods used. Observational studies using simulated

patients (i.e., actors trained to portray patients with a specific condition) tend to yield lower estimates of the rate of counseling [79]. Few states in the US require pharmacists to provide verbal counseling to patients when prescriptions are filled, with most states requiring only that pharmacists offer to counsel patients [80]. In a study using simulated patients, Svarstad and colleagues examined the rates of verbal counseling provided in 306 community pharmacies distributed across eight states in the US [81]. Risk communication, defined as providing information about at least one side effect or precaution, occurred in 17% of visits in which a prescription for amoxicillin was presented and 31% and 37% of visits in which a prescription for ibuprofen or paroxetine, respectively, was presented. Patients who filled a prescription in states with more strict regulations concerning pharmacist counseling (e.g., pharmacists required to provide face-to-face counseling) were more likely to receive risk information than patients in states with less strict regulations.

The Internet, Direct-to-Consumer Advertising, and Social Media

The availability of medication-related information via the internet, direct-to-consumer advertising, and social media has expanded dramatically since the turn of the century [82–86]. Unfortunately, the quality and accuracy of available information vary widely from source to source, and few safeguards are in place to allow consumers to evaluate the quality of information available from different sources [86,87]. Direct-to-consumer advertising (DTCA) and social media sites managed by pharmaceutical companies present special challenges. In a recent study that examined notices of violation and warning letters issued to companies by the FDA, 95% involved a branded drug website, online paid advertisement, or an online video [88]. Of the 179 violations examined, most involved the lack of risk information or the misrepresentation of benefit information. Although few countries allow DTCA, online materials travel across borders, presenting a global challenge [89].

Finally, some websites enable consumers to provide reviews of their medications, potentially opening up a Pandora's box for the spread of anecdotal information. A recent study examined over 100 000 reviews provided on the website WebMD [90]. The investigators found that in about two-thirds of the cases where differences in patient satisfaction ratings were observed for two drugs used to treat the same condition, the differences were consistent with findings reported in the published literature. However, where differences were observed, drugs with an FDA black box warning were reviewed less favorably by patients than would be expected based on the results of published studies, suggesting that these types of warnings may bias patient judgment and decision making. Clearly, more research is needed to better understand the effects of this type of internet-facilitated patient-to-patient communication.

The Future

Much of the literature on risk communication focuses on environmental risks and the risk of disease. The field of medication risk communication is still relatively young. The extent to which findings from other areas generalize to communication concerning medication risks remains unknown. Over the next few years, much will be learned as companies evaluate their REMS. For knowledge gain to be optimized, it will be important that REMS evaluation plans include a comprehensive assessment of both proximal and distal outcomes. The conceptual model depicted in Figure 39.2 may help to structure future evaluation efforts.

More basic research is also needed to assess how people process and use information about

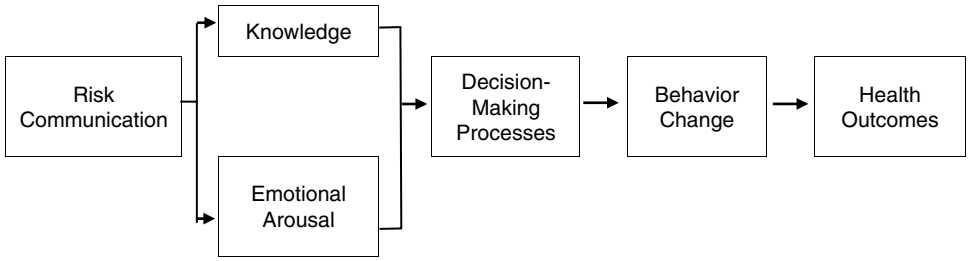


Figure 39.2 Conceptual model for evaluating the effectiveness of risk communication efforts.

medication risks. One promising approach involves the use of *fuzzy-trace theory* [91–94]. Briefly, fuzzy-trace theory posits that when an individual is exposed to risk information, two representations of the information are encoded in memory: a verbatim representation and a gist representation. The verbatim representation reflects the precise information received (e.g., 10% of patients who take Medication X experience Side Effect Y), whereas the gist representation captures the essential meaning of the information as understood by the receiver, in qualitative terms (e.g., Medication X can cause Side Effect Y). Different people exposed to the same information may form different gist representations, depending on their preexisting knowledge, previous experiences, emotional state, developmental stage, and worldview. A central tenet of fuzzy-trace theory is that when making judgments and decisions, people tend to rely on gist representations that are stored in memory and only retrieve verbatim representations when it is required by the task at hand. Further, this preference for gist processing of information increases with age and the acquisition of specialized expertise [94].

Currently, much of the risk communication literature focuses on how probabilistic information is best conveyed. From this perspective, the difficulty patients have in accurately recalling probabilistic information is viewed as problematic. However, from the perspective of fuzzy-trace theory, that conclusion might not be warranted. From a fuzzy-trace perspective, misunderstandings are most problematic when individuals interpret the gist of the information incorrectly. Numerical differences may have little effect on subsequent decisions. This possibility is supported by findings from an experimental study by Brewer *et al*. [95]. After reading a clinical vignette that portrayed a hypothetical patient, physicians in one group were asked whether the chance that the patient had a pulmonary embolism was greater or less than 1% and physicians in the other group were asked whether the chance that the patient had a pulmonary embolism was greater or less than 90%. Physicians in both groups were then asked to provide a point estimate of the chance of embolism and select from among a choice of treatment options. The irrelevant anchor (i.e., 1% versus 90%) used in the initial risk estimate had a large effect on physicians' subsequent point estimates of the probability of embolism, 23% versus 53% for physicians exposed to the low or high anchor, respectively. However, the treatment decisions made by the physicians were unaffected by the anchors. Thus, as suggested by fuzzy-trace theory, physicians appear to have based their treatment decisions on their gist representation of the information presented and were able to make rational decisions even in the presence of irrelevant information.

The findings described above illustrate the complexity of the risk communication process. Research using fuzzy-trace theory attempts to better understand the psychological processes

that underlie risk communication by systematically examining three central issues. Within the context of medication risk communication, these central issues are: (1) how do patients or clinicians extract gist from medication-related information obtained from a variety of sources (e.g., written information distributed by pharmacies when prescription medications are dispensed, direct-to-consumer advertising, healthcare providers, family/friends)? (2) what reasoning principles are invoked by contextual cues (e.g., format of the communication, images included in the communication) that affect patients' or clinicians' judgments and decisions concerning medication use? and (3) what factors (e.g., limited health literacy skills, emotional state) interfere with information processing and lead to errors in reasoning [92]? We believe that systematic research examining these types of issues has the potential to greatly expand current knowledge concerning communication of information regarding medication risks and benefits.

In conclusion, we began this chapter with the assertion that all medications have risks. The responsibility for communicating information about medication risks is shared by many entities within the healthcare system. In addition, we must recognize that we live in the Information Age. Information about medications and medication risks is disseminated by many outside the healthcare system, in some cases by individuals and groups without appropriate expertise and whose primary motive may not be the improvement of patient health outcomes. The challenge to investigators working in the field of pharmacoepidemiology is to develop communication strategies that reflect an understanding of both psychological and social issues that affect how message recipients interpret and use the information communicated.

References

1 Berry DC. *Risk, Communication and Health Psychology*. New York: Open University Press, 2004.

2 Meadows M. Promoting safe and effective drugs for 100 years. *FDA Consum* 2006; **40**(1): 14–20.

3 Lee PR, Herzstein J. International drug regulation. *Annu Rev Public Health* 1986; **7**: 217–35.

4 European Medicines Agency. What we do. www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000091.jsp&mid=WC0b01ac0580028a42 (accessed May 6, 2019).

5 Lowrance WW. *Of Acceptable Risk*. Los Altos: William Kaufmann, 1976.

6 Bogardus ST Jr, Holmboe E, Jekel JF. Perils, pitfalls, and possibilities in talking about medical risk. *JAMA* 1999; **281**(11): 1037–41.

7 Strom BL. When should one perform pharmacoepidemiology studies? In: Strom BL, ed. *Pharmacoepidemiology*, 2nd edn. New York: Wiley, 1994.

8 Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe SM, Bor DH. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* 2002; **287**(17): 2215–20.

9 Sleath B, Davis S, Sayner R, *et al*. African American patient preferences for glaucoma education. *Optom Vis Sci* 2017; **94**(4): 482–6.

10 Carpenter DM, DeVellis RF, Hogan SL, Fisher EB, DeVellis BM, Jordan JM. Use and perceived credibility of medication information sources for patients with a rare illness: differences by gender. *J Health Commun* 2011; **16**(6): 629–42.

11 Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; **294**: 218–28.

12 Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med* 2005; **353**(18): 1977–9.

13 Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA* 2007; **298**(9): 1010–22.

14 Dickinson R, Raynor DK, Knapp P, MacDonald J. How much information about the benefits of medicines is included in patient leaflets in the European Union? *A survey. Int J Pharm Pract* 2017; **25**(2): 147–58.

15 Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science* 2000; **290**(5500): 2261–2.

16 Monahan J. Statistical literacy: a prerequisite for evidence-based medicine. *Psychol Sci Public Interest* 2007; **8**(2): i–ii.

17 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007; **8**(2): 53–96.

18 Moore RA, Derry S, McQuay HJ, Paling J. What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. *Arthritis Res Ther* 2008; **10**(1): R20.

19 Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Making* 2007; **27**(5): 696–713.

20 Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: present and future research in risk communication. *Am J Health Behav* 2007; **31**(Suppl 1): S47–56.

21 Zipkin DA, Umscheid CA, Keating NL, *et al*. Evidence-based risk communication: a systematic review. *Ann Intern Med* 2014; **161**(4): 270–80.

22 Webster RK, Weinman J, Rubin GJ. People's understanding of verbal risk descriptors in patient information leaflets: a cross-sectional national survey of 18- to 65-year-olds in England. *Drug Saf* 2017; **40**(8): 743–54.

23 Buchter RB, Fechtelpeter D, Knelangen M, Ehrlich M, Waltering A. Words or numbers? Communicating risk of adverse effects in written consumer health information: a systematic review and meta-analysis. *BMC Med Inform Decis Mak* 2014; **14**: 76.

24 Peters EM, Burraston B, Mertz CK. An emotion-based model of risk perception and stigma susceptibility: cognitive appraisals of emotion, affective reactivity, worldviews, and risk perceptions in the generation of technological stigma. *Risk Anal* 2004; **24**(5): 1349–67.

25 Finucane ML. Emotion, affect, and risk communication with older adults: challenges and opportunities. *J Risk Res* 2008; **11**(8): 983–97.

26 Slovic P, Peters E, Finucane ML, Macgregor DG. Affect, risk, and decision making. *Health Psychol* 2005; **24**(4 Suppl): S35–40.

27 Slovic P, Finucane ML, Peters E, MacGregor DG. Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. *Risk Anal* 2004; **24**(2): 311–22.

28 Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA Intern Med* 2015; **175**(2): 274–86.

29 Hoffmann TC, Del Mar C. Clinicians' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA Internal Med* 2017; **177**(3): 407–19.

30 Prasad V, Ho N. Why do we continue to adopt medical practices based on pathophysiology alone when we should be insisting on clinical trials? *J Clin Epidemiol* 2014; **67**(4): 361–3.

31 DeVellis RF, Blalock SJ. Outcomes in self-management. In: Newman S, Steed E, Mulligan K, eds. *Chronic Physical Illness: Self Management and Behavioural Interventions*. New York: McGraw-Hill, 2009, pp. 148–68.

32 Reyna VF, Nelson WL, Han PK, Pignone MP. Decision making and cancer. *Am Psychol* 2015; **70**(2): 105–18.

33 Blalock SJ, Reyna VF. Using fuzzy-trace theory to understand and improve health judgments, decisions, and behaviors: a literature review. *Health Psychol* 2016; **35**(8): 781–92.

34 Raynor DK, Bryant D. European Public Assessment Report (EPAR) summaries for the public: are they fit for purpose? A user-testing study. *BMJ Open* 2013; **3**(9).

35 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module V – Risk management systems. //www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/02/WC500123208.pdf (accessed May 6, 2019).

36 Council of the European Communities. Council Directive 92/27/EEC of 31 March 1992 on the labelling of medicinal products for human use and on package leaflets. https://publications.europa.eu/en/publication-detail/-/publication/e112d726-025d-46e8-9281-3ff46198e973 (accessed May 6, 2019).

37 Dickinson D, Raynor DKT. Ask the patients – they may want to know more than you think. *BMJ* 2003; **327**(7419): 861.

38 Raynor DK, Blenkinsopp A, Knapp P, *et al*. A systematic review of quantitative and qualitative research on the role and effectiveness of written information available to patients about individual medicines. *Health Technol Assess* 2007; **11**(5): iii.

39 Berry D, Raynor D, Knapp P. Communicating risk of medication side effects: an empirical evaluation of EU recommended terminology. *Psychol Health Med* 2003; **8**(3): 251–63.

40 Berry DC, Knapp P, Raynor DK. Provision of information about drug side-effects to patients. *Lancet* 2002; **359**(9309): 853–4.

41 European Commission. A guideline on summary of product characteristics. https://ec.europa.eu/search/?query_source=PUBLICHEALTH&QueryText=https%3A%2F%2Fec.europa.eu%2Fhealth%2Fsites%2Fhealth%2Ffiles%2Ffiles%2Feudralex%2Fvol%E2%80%932%2Fc%2Fsmpc_guideline_rev2_en.pdf+&op=&swlang=en&form_build_id=form-UhlOuOzVB-8TwojSiIzZIV1gdlr4w7m07H29cF167Y8&form_id=nexteuropa_europa_search_search_form (accessed May 6, 2019).

42 Knapp P, Raynor DK, Berry DC. Comparison of two methods of presenting risk information to patients about the side effects of medicines. *Qual Saf Health Care* 2004; **13**(3): 176–80.

43 Gardner PH, McMillan B, Raynor DK, Woolf E, Knapp P. The effect of numeracy on the comprehension of information about medicines in users of a patient information website. *Patient Educ Couns* 2011; **83**(3): 398–403.

44 Knapp P, Gardner PH, Woolf E. Combined verbal and numerical expressions increase perceived risk of medicine side-effects: a randomized controlled trial of EMA recommendations. *Health Expect* 2016; **19**(2): 264–74.

45 Webster RK, Weinman J, Rubin GJ. How does the side-effect information in patient information leaflets influence people's side-effect expectations? A cross-sectional national survey of 18- to 65-year-olds in England. *Health Expect* 2017; **20**(6): 1411–20.

46 Garcia-Retamero R, Cokely ET. Designing visual aids that promote risk literacy: a systematic review of health research and evidence–based design heuristics. *Hum Factors* 2017; **59**(4): 582–627.

47 Hamrosi K, Dickinson R, Knapp P, *et al*. It's for your benefit: exploring patients' opinions about the inclusion of textual and numerical benefit information in medicine leaflets. *Int J Pharm Pract* 2013; **21**(4): 216–25.

48 Dickinson RJ. The inclusion of a headline section and information about the benefits of medicines in written medicines information (PhD Thesis). University of Leeds, 2014.

49 Slovic P, Finucane ML, Peters E, MacGregor DG. The affect heuristic. *Eur J Operation Res* 2007; **177**(3): 1333–52.

50 Blalock SJ, Sage A, Bitonti M, Patel P, Dickinson R, Knapp P. Communicating

information concerning potential medication harms and benefits: what gist do numbers convey? *Patient Educ Couns* 2016; **99**(12): 1964–70.

51 Food and Drug Administration. Guidance: Useful written consumer medication information (CMI). www.fda.gov/media/72574/download (accessed May 6, 2019).

52 Svarstad BL, Mount JK, Tabak ER. Expert and consumer evaluation of patient medication leaflets provided in U.S. pharmacies. *J Am Pharm Assoc* 2005; **45**(4): 443–51.

53 Raynor DK, Svarstad B, Knapp P, *et al*. Consumer medication information in the United States, Europe, and Australia: a comparative evaluation. *J Am Pharm Assoc* 2007; **47**(6): 717–24.

54 Winterstein AG, Linden S, Lee AE, Fernandez EM, Kimberlin CL. Evaluation of consumer medication information dispensed in retail pharmacies. *Arch Intern Med* 2010; **170**(15): 1317–24.

55 Food and Drug Administration. Format and Content of a REMS Document Guidance for Industry. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf (accessed May 6, 2019).

56 Meyer BM. The Food and Drug Administration Amendments Act of 2007: Drug safety and health-system pharmacy implications. *Am J Health Syst Pharm* 2009; **66**(Suppl 7): S3–5.

57 Food and Drug Administration. Approved Risk Evaluation and Mitigation Strategies (REMS): Prolia. www.accessdata.fda.gov/scripts/cder/rems/index.cfm?event=IndvRemsDetails.page&REMS=43 (accessed May 6, 2019).

58 Food and Drug Administration. Approved Risk Evaluation and Mitigation Strategies (REMS). www.accessdata.fda.gov/scripts/cder/rems/index.cfm (accessed May 6, 2019).

59 Food and Drug Admnistration. Medication Guides. www.fda.gov/Drugs/DrugSafety/ucm085729.htm (accessed May 6, 2019).

60 Wolf MS, King J, Wilson EA, *et al*. Usability of FDA-approved medication guides. *J Gen Intern Med* 2012; **27**(12): 1714–20.

61 Doak C, Doak L, Root J. *Assessing Suitability of Materials. Teaching Patients with Low Literacy Skills*, 2nd edn. Philadelphia: Lippincott, 1996.

62 Knox C, Hampp C, Willy M, Winterstein AG, Dal Pan G. Patient understanding of drug risks: an evaluation of medication guide assessments. *Pharmacoepidemiol Drug Saf* 2015; **24**(5): 518–25.

63 Enger C, Younus M, Petronis KR, Mo J, Gately R, Seeger JD. The effectiveness of varenicline medication guide for conveying safety information to patients: a REMS assessment survey. *Pharmacoepidemiol Drug Saf* 2013; **22**(7): 705–15.

64 Schwartz L, Mazzola N, Hoffman RS, Howland MA, Mercurio-Zappala M, Nelson LS. Evaluating patients' understanding of printed warfarin medication information. *J Pharm Pract* 2015; **28**(6): 518–22.

65 Shiffman S, Gerlach KK, Sembower MA, Rohay JM. Consumer understanding of prescription drug information: an illustration using an antidepressant medication. *Ann Pharmacother* 2011; **45**(4): 452–8.

66 Food and Drug Administration. Medication Guides – Distribution Requirements and Inclusion in Risk Evaluation and Mitigation Strategies (REMS). www.fda.gov/regulatory-information/search-fda-guidance-documents/medication-guides-distribution-requirements-and-inclusion-risk-evaluation-and-mitigation-strategies (accessed May 6, 2019).

67 Amgen. BL 125320 Prolia® (Denosumab). Risk Evaluation and Mitigation Strategy (REMS). www.fda.gov/media/78868/download (accessed May 6, 2019).

68 Amgen. Prolia (denosumab) Risk Evaluation and Mitigation Strategy. www.proliahcp.com/risk-evaluation-mitigation-strategy/ (accessed May 6, 2019).

69 Morris L. Evaluating REMS tools: issues in survey design. *Drug Inform J* 2009; **43**: 735–47.

70 Wu J, Juhaeri J. The US Food and Drug Administration's Risk Evaluation and Mitigation Strategy (REMS) Program – Current status and future direction. *Clin Ther* 2016; **38**(12): 2526–32.

71 McGrath JM. Physicians' perspectives on communicating prescription drug information. *Qual Health Res* 1999; **9**(6): 731–45.

72 Adams RJ, Appleton S, Wilson DH, Ruffin RE. Participatory decision making, asthma action plans, and use of asthma medication: a population survey. *J Asthma* 2005; **42**(8): 673–8.

73 Horne R. Compliance, adherence, and concordance: implications for asthma treatment. *Chest* 2006; **130**(1 Suppl): 65S–72S.

74 Wilson SR, Strub P, Buist AS, *et al*. Shared treatment decision making improves adherence and outcomes in poorly controlled asthma. *Am J Respir Crit Care Med* 2010; **181**(6): 566–77.

75 Sibley A, Latter S, Richard C, *et al*. Medication discussion between nurse prescribers and people with diabetes: an analysis of content and participation using MEDICODE. *J Adv Nurs* 2011; **67**(11): 2323–36.

76 Richard C, Lussier MT. Nature and frequency of exchanges on medications during primary care encounters. *Patient Educ Couns* 2006; **64**(1–3): 207–16.

77 Blalock SJ, DeVellis BM, DeVellis RF, Chewning B, Jonas BL, Sleath BL. Medication risk communication during rheumatology office visits. *Arthritis Care Res* 2015; **67**(2): 161–8.

78 Blalock SJ, Slota C, DeVellis BM, *et al*. Patient–rheumatologist communication concerning prescription medications: getting to the gist. *Arthritis Care Res* 2014; **66**(4): 542–50.

79 Puspitasari HP, Aslani P, Krass I. A review of counseling practices on prescription medicines in community pharmacies. *Res Social Adm Pharm* 2009; **5**(3): 197–210.

80 Resnik DB, Ranelli PL, Resnik SP. The conflict between ethics and business in community pharmacy: what about patient counseling? *J Bus Ethics* 2000; **28**(2): 179–86.

81 Svarstad B, Bultman, DC, Mount JK. Patient counseling provided in community pharmacies: effects of state regulation, pharmacist age, and busyness. *J Am Pharm Assoc* 2004; **44**: 22–9.

82 Agyapong VI, Nwankwo V, Bangaru R, Kirrane R. Sources of patients' knowledge of the adverse effects of psychotropic medication and the perceived influence of adverse effects on compliance among service users attending community mental health services. *J Clin Psychopharmacol* 2009; **29**(6): 565–70.

83 Thompson AE, Graydon SL. Patient-oriented methotrexate information sites on the Internet: a review of completeness, accuracy, format, reliability, credibility, and readability. *J Rheumatol* 2009; **36**(1): 41–9.

84 Narhi U, Helakorpi S. Sources of medicine information in Finland. *Health Policy* 2007; **84**(1): 51–7.

85 Carpenter DM, DeVellis RF, Fisher EB, DeVellis BM, Hogan SL, Jordan JM. The effect of conflicting medication information and physician support on medication adherence for chronically ill patients. *Patient Educ Couns* 2010; **81**(2): 169–76.

86 Randhawa AS, Babalola O, Henney Z, *et al*. A collaborative assessment among 11 pharmaceutical companies of misinformation in commonly used online drug information compendia. *Ann Pharmacother* 2016; **50**(5): 352–9.

87 Raban MZ, Tariq A, Richardson l, *et al*. Evaluation of web-based consumer medication information: content and usability of 4 Australian websites. *Interact J Med Res* 2016; **5**(3): e21.

88 Kim H. Trouble spots in online direct-to-consumer prescription drug promotion: a

content analysis of FDA warning letters. *Int J Health Policy Manag* 2015; **4**(12): 813–21.

89 Mintzes B. The tip of the iceberg of misleading online advertising comment on "Trouble spots in online direct-to-consumer prescription drug promotion: a content analysis of FDA warning letters". *Int J Health Policy Manag* 2016; **5**(5): 329–31.

90 Adusumalli S, Lee H, Hoi Q, Koo SL, Tan IB, Ng PC. Assessment of web-based consumer reviews as a resource for drug performance. *J Med Internet Res* 2015; **17**(8): e211.

91 Reyna VF, Adam MB. Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. *Risk Anal* 2003; **23**(2): 325–42.

92 Reyna VF. How people make decisions that involve risk: a dual-process approach. *Curr Dir Psychol Sci* 2004; **13**(2): 60–6.

93 Reyna VF, Adam MB, Poirier KM, LeCroy CW, Brainerd CJ. Risky decision making in childhood and adolescence: a fuzzy-trace theory approach. In: Jacobs JE, ed. *Development of Judgment and Decision Making in Children and Adolescents*. Mahwah: Lawrence Erlbaum Associates, 2005.

94 Reyna VF, Lloyd FJ. Physician decision making and cardiac risk: effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *J Exp Psychol Appl* 2006; **12**(3): 179–95.

95 Brewer NT, Chapman GB, Schwartz JA, Bergus GR. The influence of irrelevant anchors on the judgments and choices of doctors and patients. *Med Decis Making* 2007; **27**(2): 203–11.

40

Methods for Studying the Health Effects of Drug–Drug Interactions

Sean Hennessy[1], Charles E. Leonard[1], Joshua J. Gagne[2], James H. Flory[1,3], Colleen M. Brensinger[1], and Warren B. Bilker[1]

[1] *Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*
[2] *Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
[3] *Memorial Sloan Kettering Cancer Center, New York, NY, USA*

A drug–drug interaction (DDI) occurs when one or more drugs affect the pharmacokinetics (the body's effect on the drug) and/or pharmacodynamics (the drug's effect on the body) of one or more other drugs. In two-drug DDIs, the affected drug is called the **object** (or *victim*) and the affecting drug is called the **precipitant** (or *perpetrator*). The expected outcome of most hypothesized DDIs is an exaggeration of the major pharmacologic effect of the object, such as serious hypoglycemia from sulfonylurea antidiabetic agents or bleeding from anticoagulants. Other DDIs may result in reduced effectiveness of the object, such as the hypothesized reduced effectiveness of clopidogrel in lowering the risk of stroke resulting from the inhibition by proton pump inhibitors of the enzyme that converts clopidogrel to its active moiety. The precipitant of a DDI may or may not have an inherent effect on the health outcome in the absence of the object. For example, in a study of potential DDIs that involves warfarin as the object, nonsteroidal antiinflammatory drugs but not antibiotics as precipitants would be expected to increase the risk of bleeding in persons not taking warfarin.

Numerous pharmacokinetic and pharmacodynamic mechanisms are responsible for DDIs [1,2]. Because the pharmacokinetic pathways and pharmacodynamic effects of most drugs are not completely understood, it can take many years to identify, confirm, and fully understand a DDI. For example, tamoxifen and paroxetine were approved in 1977 and 1992 respectively, although it was not until 2003 that scientists identified a potential DDI between them that was hypothesized to reduce tamoxifen's effectiveness in lowering the frequency of breast cancer recurrence [3]. Although *in vitro* experiments, animal studies, and clinical trials are used to examine the effects of one drug on the pharmacokinetics of another drug, pharmacoepidemiologic studies are the principal way of studying the health effects of potential DDIs. This chapter focuses on methods for studying the health effects of potential DDIs.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Drug–drug interactions are a large and growing clinical and public health problem, especially in older adults, 40% of whom take five or more prescription drugs in a given month [4]. Although the frequency with which DDIs cause adverse health outcomes is not well studied, in older adults known DDIs are estimated to cause 13% of adverse drug events (ADEs) [5] and 5% of hospital admissions [6]. As new drugs are developed, old drugs are repurposed, and per capita drug consumption continues to rise, the clinical and public health consequences of DDIs are likely to rise correspondingly.

There are many approaches to identifying novel potential DDIs, including physiologically based pharmacokinetic models (see Chapter 2) and data mining of spontaneous reporting databases, social media posts, and healthcare data (see Chapter 27). Potential DDIs, however, may not have observable effects on health outcomes, and relatively few studies have examined the health effects of specific potential DDIs in populations. This leaves critical knowledge gaps for clinicians, patients, caregivers, editors of DDI compendia, and those who manage clinical decision support systems. Recognizing these knowledge gaps, stakeholders attending a 2009 meeting on DDIs made the conduct of additional research on the health effects of DDIs their principal recommendation [7].

False warnings about DDIs that are sent to clinicians in the context of automated messaging in the healthcare setting, such as during prescribing or dispensing, can reduce the use of valuable combinations of medications because of unsubstantiated fears that they may interact detrimentally. Further, physicians and pharmacists who are subjected to frequent alerts about apparently inconsequential potential DDIs often become desensitized to them in a phenomenon known as "alert fatigue" [8]. This provides additional importance to conducting studies of the health effects of potential DDIs, as well as studies of better ways in which the healthcare system can avoid harmful DDIs.

It seems likely that some subgroups of people are more or less susceptible to the effects of a given DDI than other subgroups (see Chapter 30). Therefore, providing information about the health effects of potential DDIs that is relevant to identifiable subgroups is an important goal of DDI research.

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

A number of methodologic problems are more prominent in pharmacoepidemiologic studies of DDIs than in those examining the effects of individual drugs. For example, pharmacoepidemiologic studies of DDIs usually require data from larger populations than those needed to study the effects of individual drugs. This is because, in any population, a small proportion of people will take any given drug, and a small proportion of those will concomitantly take the second drug of a drug–drug pair of interest. An additional problem is confounding by indication (see Chapters 3 and 43), which is regarded by many as the single biggest challenge in using the results from nonrandomized pharmacoepidemiologic studies to infer causation.

The problem of confounding by indication is more pronounced in DDI studies, which need to address confounding by the indications of two or more drugs. An additional problem is the inability of available healthcare data to validly identify all study outcomes of potential interest. In particular, while nonpharmacoepidemiologic studies of DDIs often examine serum drug

concentrations or other biomarkers such as the electrocardiographic QT interval or the international normalized ratio as outcome measures, pharmacoepidemiologic studies typically examine health outcomes such as clinically evident cardiac arrhythmias or serious bleeding. Further, some important outcomes of DDIs (e.g., serotonin syndrome) may not be validly identifiable using claims data. Thus, as in studies of the effects of individual drugs, investigators' ability to validly (and hopefully completely, or at least in a way that does not lead to biased results) ascertain outcomes that represent toxicity or lack of effectiveness using available healthcare data is essential. A further methodologic problem is that little attention has been given to optimizing pharmacoepidemiologic methods to perform well in screening for previously unanticipated associations (see Chapter 27), as may be desired when the goal is to identify hypotheses of novel potential DDIs. In addition to the need to increase the efficiency of such screening studies with regard to tasks performed by humans and by computers, screening large numbers of drug–drug pairs raises concerns that the conventionally accepted type I error rate of 5% may not be appropriate in such settings.

Currently Available Solutions

Available Research Designs for Studying the Health Effects of DDIs

Table 40.1 lists available pharmacoepidemiologic designs to study the health effects of DDIs using data derived from the provision of healthcare. The most basic and intuitive epidemiologic design is the cohort study, which compares the frequency of an outcome in different groups (i.e., cohorts) that are defined based on exposure. One possible but, as we shall see, generally unhelpful approach to assessing whether a health-affecting DDI exists is to measure the incidence rate (IR) of the adverse health outcome in four cohorts: (1) those taking the object with the precipitant (IR_{11}), (2) those taking the object without the precipitant (IR_{10}), (3) those taking the precipitant without the object (IR_{01}), and (4) those taking neither the object nor the precipitant (IR_{00}) (Table 40.1, Design 1). For DDI effects defined as a departure from multiplicity, an effect would be inferred if the following null hypothesis (H_0) were rejected:

$$H_0 : \left(IR_{11} / IR_{00}\right) = \left[\left(IR_{10} / IR_{00}\right) \times \left(IR_{01} / IR_{00}\right)\right]$$

This is to say that an effect of a DDI defined as departure from multiplicity would be inferred if the rate ratio for both-exposed vs neither-exposed were statistically different (i.e., either higher or lower) than the object-exposed vs neither-exposed rate ratio multiplied by the precipitant-exposed vs neither-exposed rate ratio. For DDI effects defined as a departure from additivity, an effect would be inferred if the following null hypothesis were rejected:

$$H_0 : \left(IR_{11} - IR_{00}\right) = \left[\left(IR_{10} - IR_{00}\right) + \left(IR_{01} - IR_{00}\right)\right]$$

This is to say that an effect of a DDI defined as a departure from additivity would be inferred if the rate difference between the both-exposed vs the neither-exposed were statistically different than the object-exposed vs neither-exposed rate difference plus the precipitant-exposed vs neither-exposed rate difference.

In practice, Design 1 is rarely if ever used to identify either multiplicative or additive effects of potential DDIs. This is because it implausibly assumes that neither the object nor the precipitant have clinical indications (i.e., reasons for taking the drug) that affect the outcome rate, or that these indications can be fully measured and controlled for. However, persons taking a given drug (whether object or precipitant) generally

Table 40.1 Pharmacoepidemiologic designs used to study health effects of potential drug–drug interactions.

Design	Relative measure of association	Key assumptions	Comments	Example
1. Cohort study examining incidence rate (IR) of the outcome in: (1) those taking the object with the precipitant (IR_{11}); (2) those taking the object without the precipitant (IR_{10}); (3) those taking the precipitant without the object (IR_{01}); and (4) those taking neither the object nor the precipitant (IR_{00})	Incidence rate ratio due to interaction (IRR_I), defined as $IRR_I = (IR_{11}/IR_{00})/[(IR_{10}/IR_{00}) \times (IR_{01}/IR_{00})]$*	No among-person unmeasured confounding by use of either object or precipitant	While this design yields the theoretically correct overall relative measure of association, the key assumption is implausible for most drug pairs	We are unaware of any published examples
2. Cohort (or case–control) study nested within person-time exposed to the object, comparing persons exposed vs unexposed to the precipitant	Incidence rate ratio (or odds ratio) associated with use of the precipitant among persons receiving the object	No among-person unmeasured confounding by use of precipitant No effect of precipitant in absence of object	Will show association if precipitant has inherent effect on outcome apart from interaction mechanism May be useful for precipitants with a chronic indication that is unlikely to be associated with outcome Use of a negative control object and/or negative control precipitant can help to assess validity of the key assumptions	Case–control study nested in person-time exposed to glyburide, examining the association between cotrimoxazole and serious hypoglycemia [9]
3. Cohort (or case–control) study nested within person-time exposed to the object, comparing person-time exposed to the precipitant vs the negative control precipitant	Incidence rate ratio (or odds ratio) associated with use of the precipitant vs control precipitant among persons receiving the object	No among-person unmeasured confounding by use of precipitant vs negative control precipitant No effect of precipitant in absence of object that is not shared by negative control precipitant No interaction between negative control precipitant and object	Preferable to Design 2 because use of a valid control precipitant reduces susceptibility to confounding by indication for the precipitant It can be difficult to know for certain that the control precipitant does not interact with the object or otherwise affect the rate of the outcome	Cohort study within person-time exposed to clopidogrel, examining the rate of ischemic stroke associated with individual proton pump inhibitors, each vs pantoprazole [13]

(*Continued*)

Table 40.1 (Continued)

Design	Relative measure of association	Key assumptions	Comments	Example
4. Cohort (or case–control) study nested within person-time exposed to either the object or the control object, comparing person-time exposed to the precipitant plus the object vs the precipitant plus the negative control object	Incidence rate ratio (or odds ratio) associated with use of the precipitant among users of the object vs use of the precipitant among users of the negative control object	No difference in direct effect of the object vs negative control object on the outcome No among-person unmeasured confounding by use of object vs negative control object No interaction between the precipitant and negative control object	Can help identify an inherent effect of the precipitant in the absence of the object	We are unaware of any published examples
5. Self-controlled case series (or case–crossover study) nested within person-time exposed to the object, comparing person-time exposed vs unexposed to the precipitant	Incidence rate ratio (or odds ratio) associated with use of the precipitant vs no exposure among persons receiving the object	No within-person unmeasured confounding by precipitant vs non-use of precipitant No effect of precipitant in absence of object	Self-controlled design inherently eliminates confounding by factors that remain constant within the individual over the study period Necessitates within-person variability in exposure to precipitant and accurate knowledge of onset and offset of exposure to precipitant For precipitants with an acute indication (e.g., antibiotics), Design 3 may be preferred if a valid control precipitant Results can be affected by secular or within-person trends in exposure to the precipitant	Case–crossover study nested within person-time exposed to warfarin examining within-person odds ratio for exposure to antimicrobial agents [16]

*IR_{11} is the incidence rate in person-time exposed to both the object and the precipitant; IR_{00} is the incidence rate in person-time exposed to neither the object nor the precipitant; IR_{10} is the incidence rate in person-time exposed to the object but not the precipitant; IR_{01} is the incidence rate in person-time exposed to the precipitant but not the object.

have an indication for that drug, while persons not taking the drug generally do not. Pharmacoepidemiologists often use the term "indication" as shorthand for denoting all the observed and unobserved factors that lead to a given patient receiving a particular medication rather than a comparator medication, or no treatment. If any aspect of this indication

(or contraindication, i.e., reason to avoid a given drug) directly affects the risk of the outcome or is otherwise associated with the outcome, then *confounding by indication* exists. Such confounding can cause the observed association to differ from the true causal effect. Confounding by indication is among the most important challenges facing pharmacoepidemiologists. Given the widespread potential for confounding by indication, it is often unrealistic to assume that the baseline rate of those taking a drug is the same as that in those not taking the drug.

If use of the precipitant in the absence of the object has no effect on the outcome, and if the precipitant is not used for an acute indication that affects or is otherwise associated with the outcome, then one can use Design 2. Design 2 is a cohort or nested case–control design that measures, within person-time exposed to the object, the incidence rate ratio of the outcome in those taking the precipitant versus in those not taking the precipitant. For example, Juurlink *et al*. used a healthcare database from older adults in Ontario to conduct a case-control study, nested within person-time exposed to glyburide [9]. Their aim was to examine the association between use vs nonuse of cotrimoxazole and serious hypoglycemia. They found that the adjusted odds ratio (OR) for the association between cotrimoxazole use and serious hypoglycemia was 6.6 (95% confidence interval [CI] 4.5–9.7). The exposure OR is the measure of association produced in case–control studies. If a case–control study uses a sampling frame known as *risk set sampling* for selection of controls, then the resulting OR is an unbiased estimator of the incidence rate ratio (IRR) that would have been produced by an analogous cohort study [10]. Risk set sampling randomly selects controls from the underlying cohort of those who were still at risk of the outcome when the corresponding case experienced the outcome.

The advantage of the nested case–control design for studies that use existing data is that it is less computationally intensive than the corresponding cohort study. Given the high computational intensity of cohort studies that account for time-varying exposures and potential confounders such as concomitant medications, this computational efficiency can be more important in studies of DDIs than in studies of individual drugs that do not account for time-varying exposures and confounders. However, when conducting nested case–control studies, care is needed in defining the time at which potential confounding variables are assessed. In a cohort study, it is intuitive and correct to assess confounding variables at baseline, before exposure has begun. In the case of studies of DDIs, exposure may be said to begin with the onset of concomitant intake of the object and precipitant. However, many nested case–control studies assess potential confounders as of the index date, often defined as the date of the outcome in cases and some corresponding date in controls. Potential confounders that are assessed after exposure can be affected by exposure. Adjusting for factors that are affected by exposure can introduce bias unless the analysis uses appropriate methods for handling time-varying confounding (see Chapter 43) [11]. Therefore, ascertaining covariates at the index date can introduce bias into nested case–control studies.

The IRR for presence vs absence of the precipitant among persons taking the object can be interpreted as the effect of a DDI (as in Table 40.1, Design 2) if there is no effect of the precipitant in persons not taking the object, and if there is no unmeasured confounding by the indication for the precipitant. Unfortunately, the assumption of no unmeasured confounding by the precipitant is often implausible. To assess its validity, investigators sometimes measure the corresponding association with a **negative control precipitant**. A negative control precipitant is a drug that is used in similar clinical circumstances as the potential precipitant under study, yet by virtue of the control precipitant's pharmacology is not believed to interact with

the object or to have an inherent effect on the outcome in the absence of the object that is not shared by the precipitant. In the setting of Design 2, the association with the negative control precipitant is used qualitatively to place into context and aid in the interpretation of the association measure for the precipitant of interest. For example, in the previously described study that measured the OR for the association between cotrimoxazole as the precipitant and serious hypoglycemia among persons receiving glyburide (the object), the investigators also examined the association with amoxicillin as a negative control precipitant. In that study, the association between amoxicillin and serious hypoglycemia (adjusted OR 1.5; 95% CI 0.8–2.9) helped provide reassurance that the association with cotrimoxazole (adjusted OR 6.6) was unlikely to be due primarily to confounding by the need for an antibiotic or a shared effect of all antibiotics [9].

To help distinguish a DDI from an inherent effect of the precipitant, one can measure the association between the precipitant and the outcome within the person-time exposed to a **negative control object**. A negative control object is a drug that is used for similar indications as the object under study, but is not believed to interact pharmacologically with the precipitant. For example, in a study of DDIs between sulfonylureas as objects and antihyperlipidemics as precipitants, Leonard *et al*. used metformin as a negative control object, which is not believed to interact with the precipitants [12]. In the setting of Design 2, the association with presence vs absence of the precipitant in users of the negative control object is used qualitatively to place into context and aid in the interpretation of the association measure of primary interest. For example, in the previously described study by Leonard *et al*. of sulfonylureas and antihyperlipidemics, the possibility of an association (although not quite statistically significant) between fenofibrate (a precipitant) and serious hypoglycemia among users of metformin (as a negative control object) suggested the possibility of an inherent hypoglycemic effect of fenofibrate in the absence of sulfonylureas [12].

Design 3 is just like Design 2 except that the association measure is the IRR (or OR) of the precipitant of interest explicitly versus the control precipitant, among persons taking the object. For example, Leonard *et al*. conducted a cohort study of persons taking clopidogrel, examining the rate of ischemic stroke among persons taking individual proton pump inhibitors, each versus pantoprazole as the negative control precipitant [13]. Pantoprazole was selected as the negative control precipitant because it is not a potent inhibitor of the enzyme responsible for activating clopidogrel (cytochrome P450 2C19) and therefore is believed to have a low potential for interacting with clopidogrel. The multiplicative interaction parameter can be produced either through performing a single regression that estimates, among those exposed to the object drug, the association between the precipitant vs the negative control precipitant; or performing one regression that estimates the association between the presence vs absence of the precipitant in those receiving the object and one that estimates the association between the presence vs absence of the negative control precipitant in those receiving the object, and then calculating the ratio of ratios and the corresponding confidence limit from these two regressions using the delta method [14]. The advantage of Design 3 over Design 2 is that it uses the association between the outcome and the control precipitant quantitatively rather than qualitatively.

Similarly, Design 4 is just like Design 2 except that the association measure is the IRR (or OR) for the precipitant of interest in those receiving the object drug of interest vs the precipitant of interest in those receiving a negative control object drug. As with Design 3, this parameter can be calculated either through a single regression or by combining the results of two

regressions using the delta method [14]. We are unaware of any published examples that have used this design.

Although use of a negative control precipitant can be a valuable strategy, there are at least three reasons why it is not a panacea for the problem of confounding by the indication for the precipitant. First, there are potential DDIs for which there is not a plausible negative control precipitant. For example, if one wanted to examine whether aspirin as the precipitant increased the risk of serious bleeding in patients receiving warfarin as the object, it would be difficult to identify a negative control precipitant that had the same set of indications as aspirin and was not believed to increase the risk of bleeding in patients taking warfarin. Second, even if there is a plausible negative control precipitant, there may still be residual unmeasured confounding between the precipitant and the negative control precipitant. For example, when amoxicillin is used as a negative control precipitant in studies examining cotrimoxazole as a potential precipitant, there may be residual confounding because amoxicillin and cotrimoxazole are not used in identical groups of patients. Third, there can be no guarantee that the negative control precipitant does not have an unknown interaction with the object or an unknown inherent effect on the outcome. This may be particularly true for older drugs, for which pharmacokinetic pathways and pharmacodynamic effects may be less well studied than for newer drugs.

Self-controlled designs include only persons who experienced the outcome, using each person as her/his own control. Such designs therefore inherently control for both measured and unmeasured potential confounding factors to the extent that such factors do not change within individual over the study period. Self-controlled designs are useful for identifying short-term effects of acute or intermittent exposures, which are often of interest in studies of DDIs. The self-controlled case series (SCCS) design is a self-controlled design that is analogous to the cohort design [15]. The case–crossover design is a self-controlled design that is analogous to the nested case–control study design [15].

Design 5 is a SCCS or case–crossover study nested within person-time exposed to the object, examining the IRR (for the SCCS design) or OR (for case–crossover design) associated with use versus nonuse of the precipitant. For a SCCS or case–crossover study to be feasible, there must be within-person variability in exposure to the precipitant while the person is taking the object. That is, a person whose entire time taking the object is either always co-exposed or never co-exposed to the precipitant will not contribute to the estimation of the drug interaction parameter, although they can contribute to the estimation of other model parameters such as time-varying confounders (if any) in analysis of a self-controlled study of the DDI. Thus, on one hand, self-controlled designs are better suited to examine DDIs involving precipitants that are taken acutely or episodically rather than chronically. On the other hand, acutely taken drugs often have acute indications that may affect the rate of the outcome, rendering the design susceptible to within-person confounding by indication.

For example, Schelleman *et al*. used the case–crossover design to examine the within-person association between use of antibiotics as precipitants and hospitalization for gastrointestinal bleeding among persons taking warfarin as the object [16]. They found that all antibiotics examined were associated with an elevated rate of bleeding, including those not believed to interact pharmacokinetically with warfarin. However, there were large differences among antibiotics. The observation that all antibiotics were associated with an increased rate of bleeding suggests either that all antibiotics share a mechanism for causing bleeding in persons taking warfarin (and possibly even in those not taking warfarin), or that the indication for antibiotics (acute infection) itself is associated

with bleeding in those taking warfarin (and possibly even in those not taking warfarin). Clinically, whether the increased bleeding risk observed during antibiotic use is due to a DDI, is a shared effect of all antibiotics, or is an inherent effect of infection may not matter as long as clinicians monitor anticoagulated patients carefully during episodes of acute infection. Thus, from a methodologic perspective, even though self-controlled designs are generally useful to study acute exposures, within-person confounding by the indication for drugs with acute indications may complicate their use for DDIs when the precipitants have acute indications.

Therefore, in the setting of acutely administered precipitants, a cohort study that quantitatively employs a negative control precipitant (Design 3) may be useful in addition to or perhaps instead of a self-controlled study (Design 5), provided that a good negative control precipitant is available. In addition, one could use a negative control precipitant in a case-case-time-control study [17] (see Chapter 43), although we are unaware of any studies that have used this design to study DDIs. Further, although self-controlled studies are generally thought of as a poor choice for studying chronically administered drugs, exposure to medications that are intended to be chronically administered is often actually intermittent because of poor persistence, incomplete adherence, or other reasons. Therefore, self-controlled designs can sometimes be useful for studying precipitants that are intended to be used chronically, although they may be vulnerable to persistent user bias [18].

One could consider performing a SCCS (or case–crossover) study nested within person-time exposed to the object, explicitly comparing person-time exposed to precipitant vs a control precipitant. This design would include only persons who took both the precipitant and the negative control precipitant while taking the object, and who experienced the outcome while taking the object plus either the precipitant or the control precipitant.

Suppose, for example, that an investigator wished to perform a self-controlled study to compare bleeding risk in warfarin users associated with concomitant use of cotrimoxazole, with amoxicillin as a negative control precipitant. A self-controlled study of this question would include only persons who experienced bleeding while treated with warfarin plus either cotrimoxazole or amoxicillin as a precipitant, and who also took the alternative precipitant at some point during warfarin therapy. Because few such persons are likely to exist even in a large population database, this design seems unlikely to be of practical use. However, one could quantitatively incorporate a negative control precipitant in a self-controlled study by fitting one regression that estimates the association with the precipitant in users of the object, fitting a second regression that estimates the association with the negative control precipitant in users of the object, and calculating the ratio of these ratios (with the corresponding confidence limits) using the delta method [14]. Similarly, one could quantitatively incorporate a negative control object using a self-controlled design by fitting one regression that estimates the association with the precipitant in users of the object, fitting a second regression that estimates the association with the precipitant in users of the negative control object, and calculating the ratio of these ratios (with the corresponding confidence limits) using the delta method [14]. For example, Han *et al*. used this approach to examine the association between numerous potential precipitants and serious hypoglycemia in users of sulfonylureas as objects, using metformin quantitatively as a negative control object [19].

As is evident from the discussion above, selection of a pharmacoepidemiologic design to study a specific potential DDI includes consideration of numerous factors including the existence of a plausible negative control precipitant and control object, the relative importance of among-person confounding versus within-person confounding, and whether the precipitant

is in real life taken acutely or intermittently versus chronically. Investigators studying a given potential DDI should consider using multiple, complementary research designs.

Outcome Assessment Methods

Many studies have used review of medical records to examine the validity and performance characteristics of algorithms to identify outcomes using administrative healthcare data [20]. Such studies usually examine outcomes that reliably result in treatment in the emergency department (ED) and/or hospital admission rather than office-based treatment. Thus, investigators studying the effects of potential DDIs on acute health outcomes usually study events that lead to ED treatment or hospitalization. Given the transition in the US from the *International Classification of Diseases*, 9th revision, clinical modification (ICD-9-CM) to ICD-10-CM that occurred on October 1, 2015, researchers using administrative data from this date or later in the US will need to examine the validity of algorithms that use ICD-10-CM codes for identifying outcomes.

As healthcare databases increasingly include laboratory values and vital signs, such measures can also be used as outcomes in DDI studies. A typical study design using such outcomes would examine change in a laboratory value from baseline when a precipitant is initiated in a person receiving an object. For example, changes in serum glucose were used to identify a possible DDI between the antidepressant paroxetine and the antihyperlipidemic pravastatin [21] and between proton pump inhibitors and metformin [22]. Compared to studies that rely on binary outcomes such as the occurrence of serious hypoglycemia, studies examining a continuous measure such as serum glucose require much smaller sample sizes and may raise fewer concerns about outcome validity, assuming that the laboratory value is accurately measured and recorded. A related limitation is that such measures are generally intermediate endpoints or biomarkers, rather than the actual clinical events that matter most to patients. In addition, handling of missing data deserves careful consideration, particularly if drug exposure affects the likelihood that providers measure or record the study endpoint.

Using a Positive Control Pair to Assess Assay Sensitivity

The use of a negative control precipitant and negative control object is discussed above, either as an explicit control group or implicitly to help assess the potential for confounding by the indication for the precipitant, or to help assess an inherent effect of the precipitant in the absence of the object. To assess the sensitivity of the pharmacoepidemiologic study to capture a known DDI similar to the one being studied (i.e., demonstrate the sensitivity of the pharmacoepidemiologic assay), investigators should consider studying a **positive control precipitant**, which is a precipitant known to produce an association with an outcome in patients receiving the object of interest. For example, if one were to study a DDI between warfarin as the object and an antibiotic as the precipitant with bleeding as the outcome, it may be useful to reproduce the well-established DDI between warfarin and cotrimoxazole as a positive control to demonstrate the ability of the study procedures and database to reproduce this known positive association. While this can be helpful, the investigator should consider the possibility that confounding might be different for the precipitant and positive control such that replicating the known association for the positive control is no guarantee that the study will yield the truth for the precipitant. In addition, other considerations such as sample size may negate the ability of a precipitant to serve as a reliable positive control.

Considering Initiation Order of Object and Precipitant

Concomitant administration of an object and a precipitant can be divided into three categories based on order of initiation of the two drugs. When both drugs are initiated simultaneously, the concomitancy is **combination triggered**. When the object is started in a person already taking the precipitant, concomitancy is **object triggered**. When the precipitant is started in a person already taking the object, concomitancy is **precipitant triggered**.

An adverse event due to a DDI involving a dose-titrated object may be more likely if concomitancy is precipitant triggered rather than either object triggered or combination triggered. This is because in precipitant-triggered concomitancy, the dose of the object may be titrated to produce its desired effect in a patient who is not receiving the precipitant, and this titration is later followed by initiation of the precipitant. For example, if warfarin is started and cotrimoxazole is later added, the prescriber may be unaware of the need to retitrate the dose of warfarin, and overanticoagulation and bleeding may result. In contrast, if warfarin and cotrimoxazole are started simultaneously or if warfarin is started in a patient already receiving cotrimoxazole, the warfarin dose will be titrated to the desired level of laboratory-measured anticoagulation in the presence of cotrimoxazole, avoiding clinical consequences of the DDI in that patient, provided that the patient continues to take cotrimoxazole. Naturally, if the cotrimoxazole is later discontinued, the patient may be at risk of the effects of underanticoagulation, that is, thromboembolic events.

If an investigator wished to include only instances of precipitant-triggered concomitancy to increase the likelihood of identifying a clinically important DDI, a larger study population would naturally be needed to detect the same level of increased risk, since only a subset of all instances of concomitancy are precipitant triggered. If sufficient sample size is available, it may be desirable to calculate separate measures of association for precipitant-triggered, object-triggered, and combination-triggered concomitancy when studying dose-titrated objects.

When studying precipitant-triggered and object-triggered concomitancy, it is critical to avoid including immortal person-time (see Chapter 43). Immortal person-time is a period of observation that is guaranteed to be event free through design of the study [23]. In an analysis of a putative DDI between clopidogrel (object) and proton pump inhibitors (precipitants), Stockl *et al.* compared clopidogrel initiators to clopidogrel initiators who also filled a prescription for a proton pump inhibitor [24]. Follow-up began at clopidogrel initiation, and patients were classified into clopidogrel-only or clopidogrel-plus-proton pump inhibitor groups based on whether they had at least one prescription for a proton pump inhibitor in the 90 days before or 90 days after the clopidogrel initiation date. Thus, patients who qualified for inclusion by receiving a proton pump inhibitor in the 90 days following the clopidogrel initiation contributed immortal person-time to the analysis – the time from clopidogrel initiation to the proton pump inhibitor prescription – since patients that entered the analysis in this way, by definition, could not have had a fatal outcome in this period. Beginning follow-up after or at the time of (but not before) concomitancy can help to avoid immortal person-time bias.

Studying the Time Course of the DDI

Even in the absence of a potentially interacting drug, the rate of an ADE often varies with amount of time since initiating the drug. This is part of the rationale for the increasingly standard practice in pharmacoepidemiology to restrict studies to new users of the drugs being examined, an approach known as the **inception cohort design** [25]. For many drug–outcome pairs, the incidence rate would

be expected to peak shortly after starting the drug and decline thereafter.

Such a declining pattern may be attributed to at least three different mechanisms. The first mechanism is depletion of susceptible patients, in which patients with an inherent susceptibility to the drug's adverse effect experience the adverse effect soon after initiation, and subsequently discontinue the drug because of the adverse event or a prodrome thereof [26,27]. Under this mechanism, the patients who remain on the drug for the long term are more robust to the drug's adverse effects, since the susceptible patients have been depleted from the cohort. The second mechanism leading to a declining event rate over time is biological adaptation to the drug's pharmacologic effects. The third mechanism is dose reduction prompted either by early signs of toxicity (e.g., a reduction in the dose of a statin due to mild myopathy that reduces the risk of rhabdomyolysis) or in response to measurement of the serum drug concentration or other biomarker used in clinical practice to adjust doses (e.g., a reduction in warfarin dose due to supratherapeutic values of the international normalized ratio, a laboratory marker of warfarin's pharmacologic effect). While each of these mechanisms would be expected to produce a declining rate, an increasing rate can be observed for drug–outcome pairs that are characterized by cumulative toxicity, such as corticosteroid-induced avascular necrosis and anthracycline-induced cardiomyopathy.

Given that the rate of an ADE often varies with the amount of time since initiating the drug, it is predictable that the rate of an outcome caused by a DDI may vary as a function of the amount of time since initiation of concomitancy, particularly for DDIs acting through metabolic inhibition [28]. The initial increase in plasma concentration of the object may cause a rise in the rate of the ADE initially, followed by a reduction in the rate as the metabolism of the object returns to baseline.

Figure 40.1A illustrates a scenario in which initiation of a precipitant to a person already receiving an object (i.e., precipitant-triggered concomitancy) leads to an event rate that is transiently increased but then declines to baseline. The rate might actually decline to below the baseline rate because the persons susceptible to the adverse effect become depleted from the cohort or because the body compensates to increase pharmacologic clearance of the object. If the scenario illustrated in Figure 40.1A is operating, and one evaluates a potential DDI by calculating the average rate during all time treated with the object–precipitant combination and dividing this rate by the rate observed during the time treated with the object alone, then one could falsely conclude that the potential DDI had no effect on the rate of the adverse event, even if the precipitant has a large but transient effect. This is because, as illustrated in Figure 40.1A, the transiently increased rate seen shortly after the initiation of the precipitant in patients is outweighed by the prolonged time during which the rate of the adverse event has reverted back to (or even below) the baseline rate associated with use of the object alone. In other scenarios, the increased risk associated with a precipitant-triggered DDI may remain elevated throughout the course of concomitancy, as illustrated in Figure 40.1B.

Careful consideration must also be given to the timing of concomitancy when the rate of the ADE varies with the amount of time since initiating the object. For example, a study of a DDI between corticosteroids and some precipitant on avascular necrosis should account for time on corticosteroids since the rate of avascular necrosis increases with time on corticosteroids. If, for example, an investigator conducted an analysis in which a large portion of the time unexposed to the precipitant was shortly after corticosteroid initiation, and the majority of time concomitantly exposed to the precipitant was longer after corticosteroid initiation, then there would be a lower baseline risk of avascular necrosis during unexposed time than in exposed time, even if there were no effect of the precipitant.

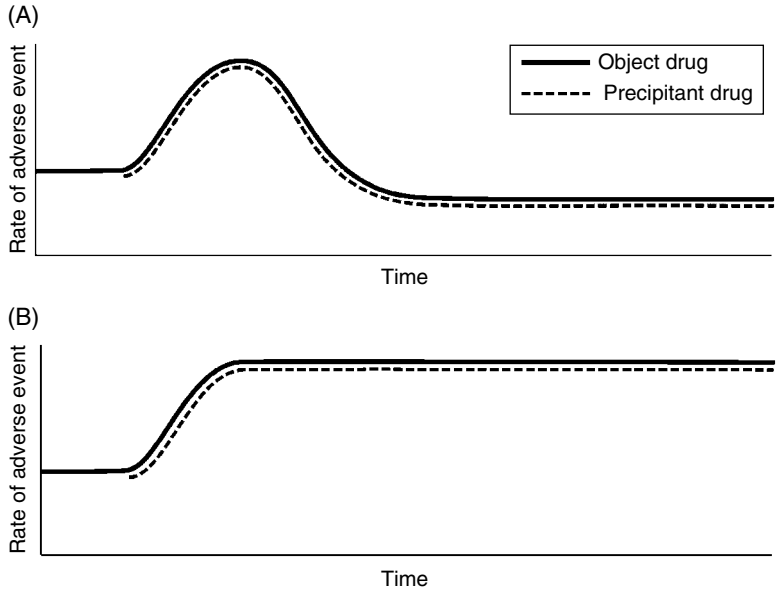


Figure 40.1 Schematic depiction of the two different potential time-courses of a precipitant-triggered drug–drug interaction. In (A), the rate of the adverse event rises transiently, while in (B), the rate of the adverse event rises and remains persistently elevated.

Given that, as illustrated in Figure 40.1A, the *overall* rate ratio may not be observably elevated for a DDI with a substantial but transient effect, it can be important to look for an association within time-specific strata (i.e., examine the duration–response relationship of the DDI) regardless of whether or not an overall association is observed over the entire period of concomitancy. However, looking for associations both overall and within strata defined by time since initiation of concomitancy can raise potential concerns about multiple testing. Given that DDI studies using even very large population databases can have low statistical power, adjusting for multiple comparisons across time strata may have a crippling effect on investigators' ability to identify important risks associated with DDIs. For example, Schelleman *et al.* used a population database of approximately 108 million person-years of follow-up to evaluate potential DDIs involving sulfonylureas as objects and lipid-lowering drugs as precipitants [29]. They studied only precipitant-triggered instances of concomitancy. For each object–precipitant pair, they examined the overall association as well as associations for 0–29 days, 30–59 days, 60–119 days, and ≥120 days. They found statistically elevated association measures for several time-specific strata, but would not have done so had they accounted for multiple testing due to the 56 possible duration-specific association measures, many of which had insufficient data even to estimate a multiplicity-unadjusted measure of association.

We believe that for exploratory analyses of time-specific measures of association, refraining from accounting for multiplicity is justified because of the manifestly low statistical power associated with multiplicity-adjusted estimates, provided that such association measures are interpreted as exploratory in light of their corresponding higher-than-nominal type I error rate. The issue of multiple testing can be mitigated in settings where the pharmacologic mechanism of

the potential DDI is sufficiently well characterized so that the time course of the interaction can confidently be predicted *a priori*, and analyses within specific time windows considered primary, with other time windows considered secondary.

Pharmacoepidemiologic Screening to Identify Potential DDIs

In addition to performing hypothesis-driven DDI studies, pharmacoepidemiologic methods can be used to perform hypothesis-free screening of healthcare data to identify potential DDIs. For example, Han *et al*. used the SCCS design to screen healthcare data for precipitants that are associated with serious hypoglycemia in persons taking insulin secretagogues (Design 5) [19]. The SCCS design is well suited to screening because it includes only persons who experienced the outcome while taking the object. This makes this design highly computationally efficient and thus more amenable to high-throughput analysis than the cohort or nested case–control designs. Because of the large number of candidate precipitants that they examined, the investigators used a semi-Bayesian shrinkage approach for multiple comparisons adjustment [30], an approach that limits the variability of the resulting measures of association and controls the type I error rate.

The Future

Given the continued development of new drugs, repurposing of old drugs, the rising frequency of polypharmacy, and the aging of the population, the clinical and public health importance of DDIs will continue to grow. The increasing use of healthcare data from larger populations, including data accessed using distributed data models (see Chapter 25), that characterizes pharmacoepidemiology in general promises to be particularly important for studying the health effect of DDIs. This is because studying the effects of multiple drugs in combination necessitates larger population databases than does studying the effects of individual drugs. The settings in which the health effects of DDIs are characterized are likely to expand from the current predominance of studies of community-dwelling persons to those set in hospitals, nursing homes, and other settings.

A wide variety of data and approaches are now being used to screen for potentially clinically important DDIs, including animal models, healthcare data, spontaneous reporting data (see Chapter 10), physiologically based pharmacokinetic models (see Chapter 2), physiologic and pharmacologic networks (see Chapter 2), and social media (see Chapter 27). As the use of screening increases, the number of hypothesized DDIs whose health effects need to be confirmed or refuted in etiologic studies will rise. Perhaps the most urgent need is to develop and test approaches to better incorporate the knowledge gained through studies of the health effects of DDIs into the healthcare system, thereby reducing the frequency of harmful effects of DDIs while allowing and perhaps even encouraging use of combinations that had been predicted to be harmful but were actually found to be safe. However, given the fragmented market for DDI knowledge bases and the surprising degree of lack of agreement among them [31], addressing this problem will be challenging.

Acknowledgments

The authors thank Mary A. Leonard for preparing Figure 40.1 and Maria Kalai for assistance with manuscript preparation. Writing of this chapter was supported by grants R01AG025152 and R01DK102694 from the US National Institutes of Health and by grants R01HS023122 and K08HS023898 from the US Agency for Health Care Research and Quality.

Portions of this chapter have been previously published as Hennessy S, Leonard CE, Gagne JJ, Flory JH, Han X, Brensinger CM, Bilker WB. Pharmacoepidemiologic methods for studying the health effects of drug–drug interactions. *Clinical Pharmacology & Therapeutics* 2016; **99**(1): 92–100. Copyright © 2015, American Society for Clinical Pharmacology and Therapeutics, John Wiley & Sons (http://onlinelibrary.wiley.com/wol1/doi/10.1002/cpt.277/full)

References

1 Jia J, Zhu F, Ma X, Cao Z, Li Y, Chen YZ. Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009; **8**(2): 111–28.
2 Delafuente JC. Understanding and preventing drug interactions in elderly patients. *Crit Rev Oncol Hematol* 2003; **48**(2): 133–43.
3 Stearns V, Johnson MD, Rae JM, *et al*. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J Natl Cancer Inst* 2003; **95**(23): 1758–64.
4 Kit BK, Ogden CL, Flegal KM. Prescription medication use among normal weight, overweight, and obese adults, united states, 2005–2008. *Ann Epidemiol* 2012; **22**(2): 112–19.
5 Gurwitz JH, Field TS, Harrold LR, *et al*. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003; **289**(9): 1107–16.
6 Becker ML, Kallewaard M, Caspers PW, Visser LE, Leufkens HG, Stricker BH. Hospitalisations and emergency department visits due to drug–drug interactions: a literature review. *Pharmacoepidemiol Drug Saf* 2007; **16**(6): 641–51.
7 Hines LE, Malone DC, Murphy JE. Recommendations for generating, evaluating, and implementing drug–drug interaction evidence. *Pharmacotherapy* 2012; **32**(4): 304–13.
8 Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. *Health Aff* 2011; **30**(12): 2310–17.
9 Juurlink DN, Mamdani M, Kopp A, Laupacis A, Redelmeier DA. Drug–drug interactions among elderly patients hospitalized for drug toxicity. *JAMA* 2003; **289**(13): 1652–8.
10 Greenland S, Thomas DC. On the need for the rare disease assumption in case–control studies. *Am J Epidemiol* 1982; **116**(3): 547–53.
11 Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**(5): 550–60.
12 Leonard C, Bilker W, Brensinger C, *et al*. Severe hypoglycemia in users of sulfonylurea antidiabetic agents and antihyperlipidemics. *Clin Pharmacol Ther* 2016; **99**(5): 538–47.
13 Leonard CE, Bilker WB, Brensinger CM, *et al*. Comparative risk of ischemic stroke among users of clopidogrel together with individual proton pump inhibitors. *Stroke* 2015; **46**(3): 722–31.
14 Bieler GS, Williams RL. Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics* 1993; **49**(3): 793–801.
15 Maclure M, Fireman B, Nelson JC, *et al*. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 50–61.
16 Schelleman H, Bilker WB, Brensinger CM, Han X, Kimmel SE, Hennessy S. Warfarin with fluoroquinolones, sulfonamides, or azole antifungals: interactions and the risk of hospitalization for gastrointestinal bleeding. *Clin Pharmacol Ther* 2008; **84**(5): 581–8.

17 Wang SV, Gagne JJ, Glynn RJ, Schneeweiss S. Case-crossover studies of therapeutics: design approaches to addressing time-varying prognosis in elderly populations. *Epidemiology* 2013; **24**(3): 375–8.

18 Hallas J, Pottegård A, Wang S, Schneeweiss S, Gagne JJ. Persistent user bias in case-crossover studies in pharmacoepidemiology. *Am J Epidemiol* 2016; **184**(10): 761–9.

19 Han X, Chiang CW, Leonard CE, Bilker WB, Brensinger CM, Hennessy S. Biomedical informatics approaches to identifying drug–drug interactions: application to insulin secretagogues. *Epidemiology* 2017; **28**(3): 459–68.

20 Carnahan RM. Mini-sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 90–9.

21 Tatonetti NP, Denny JC, Murphy SN, *et al*. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin Pharmacol Ther* 2011; **90**(1): 133–42.

22 Flory J, Haynes K, Leonard CE, Hennessy S. Proton pump inhibitors do not impair the effectiveness of metformin in patients with diabetes. *Br J Clin Pharmacol*.2015; **79**(2): 330–6.

23 Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf* 2007; **16**(3): 241–9.

24 Stockl KM, Le L, Zakharyan A, *et al*. Risk of rehospitalization for patients using clopidogrel with a proton pump inhibitor. *Arch Intern Med* 2010; **170**(8): 704–10.

25 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**(9): 915–20.

26 Moride Y, Abenhaim L. Evidence of the depletion of susceptibles effect in non–experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**(7): 731–7.

27 Guess HA. Behavior of the exposure odds ratio in a case–control study when the hazard function is not constant over time. *J Clin Epidemiol* 1989; **42**(12): 1179–84.

28 Amin M, Suksomboon N. Pharmacotherapy of type 2 diabetes mellitus: an update on drug–drug interactions. *Drug Saf* 2014; **37**(11): 903–19.

29 Schelleman H, Han X, Brensinger CM, *et al*. Pharmacoepidemiologic and in vitro evaluation of potential drug–drug interactions of sulfonylureas with fibrates and statins. *Br J Clin Pharmacol* 2014; **78**(3): 639–48.

30 Steenland K, Bray I, Greenland S, Boffetta P. Empirical bayes adjustments for multiple results in hypothesis-generating or surveillance studies. *Cancer Epidemiol Biomarkers Prev* 2000; **9**(9): 895–903.

31 Abarca J, Malone DC, Armstrong EP, *et al*. Concordance of severity ratings provided in four drug interaction compendia. *J Am Pharm Assoc* 2004; **44**(2): 136–41.

41

The Pharmacoepidemiology of Medication Errors

Hanna M. Seidling[1,2] *and David W. Bates*[1]

[1] *Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA*
[2] *Department of Clinical Pharmacology and Pharmacoepidemiology, Cooperation Unit Clinical Pharmacy, University Hospital Heidelberg, Heidelberg, Germany*

Medications represent the most commonly used form of medical therapy today. For adults, 75% of office visits to general practitioners and internists are associated with the continuation or initiation of a drug [1]. For hospitalized patients, multiple medication orders tend to be written for each patient daily. Theoretically, medication errors can refer to selection of the wrong patient, the wrong drug, the wrong galenic formulation (e.g., tablets with immediate and sustained release), the wrong dosage or route of administration, or wrong time. Medication errors are frequent, but fortunately only a small proportion result in harm [2]. However, given the high prevalence of prescription medication use, preventable adverse drug events are one of the most frequent causes of preventable iatrogenic injuries. The IoM report "To Err is Human" suggested that at least 44 000–98 000 deaths occur in the US from iatrogenic injury [3]. One study estimated that about 7000 deaths are attributed to medication errors [4] and about 1 million injuries might result from medication use in general in the US per year.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Definition and Classification of Medication Errors

While the techniques of pharmacoepidemiology have most often been used to study the risks and benefits of drugs, they can also be used to study medication errors and their attendant adverse drug events. Medication errors have been defined as "any error in the process of ordering, dispensing, or administering a drug" regardless of whether an injury occurred or the potential for injury was present [5]. Mechanistically, medication errors may result from errors in planning actions – for example, not knowing the correct starting dosage for a medication (i.e., knowledge-based mistakes or rule-based mistakes) – or errors in executing correctly planned actions, like picking one sound-alike medication instead of another (i.e., action-based slips or memory-based lapses) [6].

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

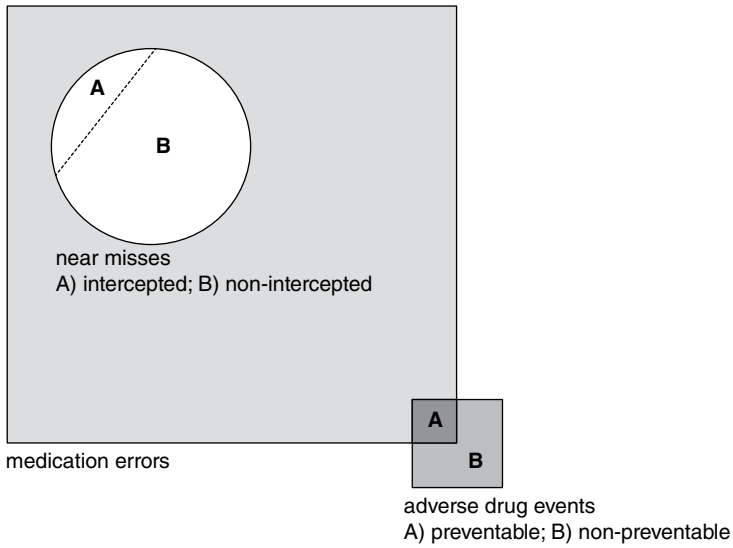


Figure 41.1 Relationship of medication errors and adverse drug events. About 1 in 10 medication errors is likely to result in patient harm [7], whereas about 25% of adverse drug events can be allocated to a medication error [2]. Near misses, both intercepted and nonintercepted, comprise those medication errors with potential for patient harm without resulting in actual harm.

In clinical practice, a medication error may occur at any stage of drug therapy, including drug prescribing, transcribing, manufacturing, dispensing, administering, and monitoring. Medication errors with potential for harm are called near-misses or potential adverse drug events; these errors may be intercepted before they reach the patient, or reach the patient without consequence. However, generally, about one in 10 medication errors results in patient harm [7]. An adverse drug event (ADE) would be considered preventable if a medication error is associated with the ADE (Figure 41.1). While ADEs have been defined as "any injury related to the use of the drug, regardless of whether a therapeutically appropriate dosage is used, although the causality of this relationship may not be proven" [8], an adverse drug reaction (ADR) can be defined as harm which is caused by a drug while appropriately used [9] (see also Chapter 1 for alternative definitions).

Detection of Medication Errors

Approaches for detecting medication errors include manual or automatic screening of claims data, administrative databases, medical records, electronic health records, or incident reports mostly by providers in hospitals, as well as patient monitoring or direct observation often by pharmacists. All approaches have inherent advantages and pitfalls and there is no single approach that is considered the gold standard for detecting medication errors or ADEs. Factors which might influence the identification of medication errors and ADEs include the setting (ambulatory vs inpatients; routine care vs research studies), the expected types of medication errors (prescribing vs administration errors), and the projected costs of detection [10]. In addition, the type of detection method influences which types of medication errors are found (e.g., only those resulting in patient harm) and with which frequency (see Chapters 8 and 10 for further discussions of detecting medication adverse events).

Screening of claims data, administrative databases, medical records, and electronic health records is used to evaluate large datasets, but is generally done retrospectively. The quality of the available information, however, varies between different data sources which limits

opportunities to comprehensively and systematically detect medication errors. Especially in the outpatient setting, claims data can be obtained for very large numbers of individuals. In the US, this represents tens and sometimes hundreds of millions of people, and in many other countries complete data for a population (such as the province of Ontario) may be available. Limitations of using claims data to detect medication errors include uncertainty regarding medication consumption and mischaracterization of the error if not linked to other information sources because clinical detail is often minimal (e.g., information on weight, actual drug dose or renal function might be missing). Since the focus of such data systems is on clinical outcomes and treatment, medication errors will be missed unless they result in patient injury severe enough to come to medical attention. Even then, it is usually not clear whether the injury was due to an error.

In the inpatient setting, manual chart review is a well-established method to detect ADEs and medication errors. With most relevant patient information at hand, the appropriateness of drug prescribing and administration can be assessed, although documentation may still be incomplete, especially for assessing issues such as appropriateness of the medication order. The main limitations with chart reviews are that they are time-consuming and expensive, with the average chart review costing approximately $20 per chart.

If electronic health records are available, the manual screening of paper-based information can be replaced by semi-automated approaches. However, the level of standardization and the extent to which clinical information is stored using controlled vocabulary determine the feasibility and effectiveness of automated, algorithm-based data analyses [11]. If electronic health records include electronic prescribing applications with clinical decision support (i.e., computerized physician order entry – CPOE), data from these applications can readily be used to detect many types of medication errors at the stage of prescribing. However, the specificity of the systems will also depend on the availability of information accessible via the electronic health records [12]. Specific types include overly high dosage, cumulative dose errors, and drug–drug interaction issues, among others.

Screening of incident reports (i.e., reports usually issued by personnel involved in the occurrence of an adverse event or a situation that might have led to an undesirable outcome) and *patient monitoring* (e.g., for specific symptoms) can each reveal medication errors that resulted in patient harm [13]. Screening of incident reports always grossly underestimates the incidence of errors (because of underreporting of events), but is relatively inexpensive because data are collected as a byproduct of routine care delivery. The major barrier for reporting medication errors is staff perception that reporting might be associated with disciplinary actions [14], even if the hospital pursues a nonpunitive policy [15]. One approach to increase reporting would be to only report near misses that help to identify which situations facilitate errors but also which actions might help to detect and prevent errors. These "critical incident reporting systems" typically belong to quality management systems in hospitals and are becoming more prevalent in the primary care setting. Predetermined patient monitoring for adverse drug events, while more time- and cost-intensive, has been successful, and can identify more adverse drug events than chart review [13].

Spontaneous reporting of medication errors (described in Chapter 10) is comparatively easy to implement and to maintain, both in inpatient and outpatient settings. However, both ADEs and medication errors are substantially underreported (see Chapter 10). Nevertheless, spontaneous reporting is useful for obtaining samples of errors. However, this method cannot be used to assess the underlying rate of medication errors in a sample [16].

Direct observation is typically conducted during research studies and offers a comprehensive assessment of medication dispensing and administration errors. While being both cost- and personnel-intensive, direct observation has been successfully and reliably used to classify complex medication errors [17], and is particularly useful at stages that are not sensitive to other detection methods (e.g., drug preparation or drug administration) [18].

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

Pitfalls in the Detection of Medication Errors

The reliable and systematic detection of medication errors has many methodologic challenges, including the definition of what constitutes a medication error and the availability and appropriate interpretation of clinical data.

With respect to definition, examples of complexities include whether there was harm or potential for harm, and the decision about whether to include errors that are intercepted before reaching the patient.

Identification of medication errors remains challenging as general standards are lacking. For instance, the detection of "wrong timing errors" (i.e., giving a drug within a timeframe) requires the definition of a threshold value above which the medication is delayed. In the inpatient setting, this threshold value might be two or four hours, depending on the institution. However, sometimes patients are away from their inpatient rooms (e.g., getting diagnostic tests), in which case decisions need to be made about whether to use a singular threshold value.

Using the example of hazardous prescription of interacting drugs, a potential approach to detect a medication error involves the comparison of the prescribed medications with a drug–drug interaction (DDI) knowledge base. However, the content of such knowledge bases varies widely, in terms of both included drug pairs and specific information linked to a drug pair (e.g., severity of the DDI) [19] – see Chapter 40. Especially in the outpatient setting, comprehensive and reliable data on the patient's medication list may be missing. Furthermore, prescribing and dispensing data are seldom jointly available and determining actual patient adherence is even more difficult. Even patient surveys may not give adequate information. While patients might be nonadherent to some prescribed drugs, they might also consume over-the-counter drugs with potential for DDIs (e.g., St John's wort) that they do not report [20].

To evaluate the appropriateness of a medication for a specific patient, knowledge of the patient's characteristics is mandatory. For example, many medications are contraindicated in pregnancy, with notable examples being thalidomide, isotretinoin, and warfarin. In this context, the greatest difficulty lies in assessing whether the patient is pregnant at the time of the exposure. Information on whether a woman is pregnant or not at the time of prescribing is challenging to obtain and most information systems do not have good approaches for tracking this. In retrospective analyses, identification of the date of birth and backward calculation under the assumption of a term pregnancy might be feasible, though this process can still be subject to misclassification (e.g., if the pregnancy was not full term) and can be complex since such information is not readily stored in one location.

Another important piece of clinical information, especially in pediatrics (though also for the administration of chemotherapy and some other situations), is the patient's weight. Most pediatric medications use weight-based dosing. Standardized documentation of this information can be challenging to obtain, hindering not only analyses of pediatric dosing but also actual prescribing by pediatricians. Obtaining accurate

weight is also essential for many oncology patients, as certain intravenous chemotherapy drugs use weight-based dosing. However, this issue is further complicated in obese patients who may require dosing using body surface area (BSA) or ideal body weight (IBW).

Finally, information on the patient's medication allergy status is infrequently and inconsistently available [21,22]. It is important that true allergies (e.g., a rash related to penicillin) be differentiated from medication sensitivities or intolerances (nausea from codeine) through coded information rather than free text. It is particularly important that severe reactions, such as anaphylaxis, are clearly coded and identifiable. The eventual aim is to have one universal medication allergy list in an electronic format for each patient, rather than multiple disparate lists.

Measuring Incidence of Medication Errors

Especially because of the different approaches used in detection of medication errors, the assessment of medication error incidence remains challenging. Comparison of medication error incidence rates among different studies has substantial limitations. This is related to disparate detection approaches and using different methods to ascertain numerators (i.e., the medication errors) and denominators (i.e., the sample from which the medication errors arise). Thus, medication error rates from different studies can be difficult to compare unless the same, or similar, methods were used. Other factors to consider are the setting studied and the patient population. In addition, spontaneous reporting typically lacks information to calculate incidence (see Chapter 10).

Comparing Medication Error Rates Across Settings

Most medication error and ADE studies have been performed in the *hospital* setting. In the inpatient adult setting, patients are vulnerable to medication errors due to their medical acuity, the complexity of their disease process and medication regimens, and their age (e.g., the elderly are particularly susceptible). The medication error rate may differ depending on the type of hospital and may be higher in nonuniversity hospitals. A review from 2007 indicates that medication errors occur in about 5.1% (range 0.038–26%) of medications dispensed in university hospitals and 13.7% (range 3.5–49%) in nonuniversity hospitals [7]. Studies of ADE rates in hospitals have found rates ranging from 2 to 15 per 100 admissions [5,23,24].

In *intensive care units* (ICUs), the rates of medication errors appear higher than on general care units. This may result from the administration or ordering of many more medications that may also be associated with higher levels of toxicity. Beyond the increased incidence of medication errors in ICUs, the nature and causes of medication errors are different and the risk that a medication error will result in patient harm is also higher compared to general inpatient wards [25], with 7.4% of patients experiencing an ADE resulting from a medication error [26].

In *nursing homes* and especially in the *ambulatory setting* [13], assessment of medication error incidence is challenging because the individual steps in the medication process are rarely jointly documented (e.g., administration), and there are often substantial time lags between them. Sometimes estimation of frequency of medication errors has relied on spontaneous reporting of medication errors [27] or documentation of ADEs in charts, which misses both many ADEs and nearly all medication errors. In a recent review of medication errors in nursing homes, medication errors were reported for 16–27% of the residents. Most errors were associated with mild effects and only 0–1% resulted in severe effects [28].

In the ambulatory setting, patients live in their homes and take their medications independently,

which makes detection of medication errors and ADEs challenging. In one review, medication error rates ranged from 12% to 59%, with even higher numbers in elderly patients with complex medication regimes [29]. Thereby, errors can also be committed by a third person such as a caregiver who is also responsible for drug administration [30]. In addition, the incidence of medication error-related ADEs may be estimated by direct patient surveys in the outpatient setting, for example by calling patients or mailing or emailing them a survey. Using this kind of approach, ADE rates ranged from 25% of patients (as self-reported in a survey) [13] to 5% (of hospital admissions) [31]. For medication errors related to the prescription process, the error rate was 7.6% of all prescription orders in one study [32]. Medication error rates stratified for different specializations or dentists have not been studied in detail [33].

Another issue is what happens at the interfaces of care, for example when a patient is discharged home from the hospital. Many studies have shown that discrepancies in drug treatment at transitions of care are frequent and often these discrepancies are unintentional, facilitating substantial risk for patients [34,35]. For example, at the interface between primary and tertiary care [36], and especially in the elderly population, the incidence of problems with the drug prescription regime are frequent after discharge (in about one-third of elderly, discharged patients) and contribute to higher rehospitalization rates [37].

Comparing Medication Error Rates Across Different Patient Populations

Most early studies on medication errors and ADE have been done in *adults*. Medication errors were common, occurring at a rate of 5 per 100 medication orders in inpatients [2]. Seven in 100 medication errors had significant potential for harm, and 1 in 100 actually resulted in an injury [2].

In primarily the inpatient setting, medication error rates in *pediatric patients* have been estimated to be as high as 5–27% of all medication orders [38]. In neonatal intensive care units, error rates have been reported to be in similar ranges [39]. In the outpatient setting in cancer patients, medication error rates were three times higher in pediatric patients (18.8% of patients) than in adult patients (7.1% of patients) [40].

Medication error-related ADE rates have also been reported for the elderly; as many as 35% of elderly outpatients per year may experience an ADE [41], and as much as 30% of hospital admissions are ADE related in the elderly [42]. In elderly patients, many medication error studies have focused on the prescription of inappropriate drugs, especially using the Beers criteria (a list of drugs specified through expert consensus that should be avoided in elderly patients in general or under consideration of specific cofactors including co-morbidity or dosage) [43],although the utility of these criteria has been challenged [44].

Comparing Medication Error Rates Across Detection Methods

The incidence of medication errors may vary as much as 100-fold depending on the detection method. While direct observation is the most cost-intensive approach (about $5 per evaluated medication), it will yield the most accurate estimation of medication error incidence for dispensing and administration errors [45]. When aiming to detect the same set of medication errors by chart review or incident report review, costs substantially decrease but so do numbers of detected events, from the actual incidence rate of 11.7% (direct observation) to 0.7% (chart review) and 0.04% (incident report review). Moreover, the reported incidence will depend on the training and profession of the person who conducts the detection [45].

Medication error incidence rates are grossly underestimated if voluntary reporting methods

are applied [46]. To promote reporting, non-punishment policies as well as anonymous reporting have been established. Moreover, it is especially crucial to invite all individuals in the healthcare system who might be confronted with a medication error to report the error. For example, in the outpatient setting, where patients tend to see several physicians but get their medications from a single pharmacy, medication errors may be discovered in the pharmacy rather than during doctor's consultation. Thus, pharmacists should be invited to report medication errors to improve the systematic collection methodology [47].

Measuring Impact on Health-Related Outcome

As noted earlier, in one study 7 in 100 medication errors had significant potential for harm, and 1 in 100 actually resulted in an injury [2]. More recent literature indicates that in hospitalized patients, even 1 in 10 medication errors might result in an ADE [7]. However, the risk of whether a medication error results in harm varies. For example, the susceptibility to suffer an ADE is higher in geriatric wards as well as ICU patients compared to general care units (12% vs 6%) [25]. On the other hand, in one study pediatric patients had similar rates of ADEs compared to adults but a threefold higher rate of near misses [48]. Incidence rates of ADE in hospitalized patients are reported with a median overall frequency of 6.1% of patients [7]. Again, the detection method used substantially influences the estimation of the incidence, with highest numbers found by patient monitoring [7]. In about 2.9% (range 0.14–5%) of the patients experiencing an ADE, the ADE was fatal [7]. Nonfatal ADEs might prolong the hospital stay or increase the risk of rehospitalization. In another study, 13% of patients experienced an ADE after discharge, and of these 24% were preventable and 38% ameliorable [49]. In addition, ADEs occurring in the outpatient setting can contribute to hospital admissions, with 4.5 preventable ADEs per 1000 person-months [50].

Identifying Risk Factors

The search for risk factors for medication errors has been challenging, as some appear to occur relatively randomly in the medication process. Robust systems need to detect and prevent even errors occurring randomly [51]. Substantial research has been conducted on error nascence (i.e., the origin of the medication error) and it is important to understand and acknowledge the underlying causes on system and workflow or process level facilitating error nascence [52].

To subsequently assess these causes, it can be helpful to determine:

- at what stage of the treatment process medication errors are occurring
- by which person involved in the treatment process (e.g., the physician, the nurse, the pharmacist, the patient, or an informal care person) the error might be committed or potentially intercepted
- what the patient's characteristics are, including age, co-morbidities, and other medications they are taking
- what the clinical setting is.

These factors can be grouped into the categories of system level, patient level, and medication characteristics.

Well-defined factors influencing the risk for medication errors on a *system level* include organization policies or the general safety culture of an institution. On a workflow or process level, risk factors comprise poor communication, heavy workload or inadequate procedures [52]. Also *patient factors* such as renal dysfunction and old age increase error risk. Setting is also important, ICU patients having an especially high risk, because they are more seriously ill and are exposed to large numbers of medications. Settings with only limited monitoring options such as home care appear also risky [53].

In all settings, being administered the wrong dose is the most frequent type of medication error, especially overdosage [54]. Dosage errors may occur at the stage of administration (e.g., accidental intake of two tablets), the stage of manufacturing or dispensing (e.g., misreading the brand name), or, most frequently, at the stage of prescription. To select the appropriate dose for each patient, the physician has to consider a number of *patient characteristics* (age, weight) as well as *drug characteristics*. The individual exposure to a drug is subject to changes in the elimination organ function (e.g., renal or liver disease), pharmacokinetic interacting comedication, and genetic polymorphisms. Moreover, required dosages will depend on age-related pharmacodynamic changes and vary between disease conditions. They might also be higher or lower both at the beginning or the end of the therapy. The physician needs to have all such information at hand once he/she decides to prescribe a certain drug for a specific patient – and a lack of information might result in underdosage or, more often, in overdosage.

Any drug or drug formulation can be associated with a medication error. However, there are *medication characteristics*, including active ingredients, that are associated with an increased risk for medication errors. Predisposing factors include:

- a sophisticated way of prescribing (e.g., complex dosage adjustments), administration (e.g., usage of administration devices), or monitoring (e.g., therapeutic drug monitoring)
- a substantial dose-dependent toxicity which increases the likelihood that a medication error will result in patient harm
- a prescription frequency which is high enough that the error will occur during the study period but low enough that detection can be challenging.

The drug class with the highest prescription frequency is cardiovascular drugs. Consistent with the prevalence of prescribing, cardiovascular drugs have often been associated with an increased risk of medication errors and ADEs [55]. The prescription of antibiotics also has often resulted in ADEs, most often because known allergies were ignored [55]. Medication errors with fatal outcomes, however, are often associated with drugs which are less frequently used but complicated in their mode of administration. For instance, accidental intrathecal injection of vincristine has caused many deaths [56] despite extensive error prevention measures [57]. Similarly, intravenous administration of amphotericin B is complex and carries a high risk of harm; for intravenous administration, amphotericin is used both in an aqueous and a liposomal drug formulation with 3–4-fold higher maximum recommended doses for the liposomal preparation. Erroneous administration of aqueous amphotericin B solution in dosages appropriate only for the liposomal preparation has resulted in a number of cases of renal toxicity and death [58].

Most often, drugs frequently reported in medication error studies have more than one predisposing factor. Examples include warfarin, for which treatment must be closely monitored by adapting dosages to measured INR values to maintain effectiveness and prevent ADEs such as bleeding. In one inpatient study [59], about 30% of reported ADEs were caused by inappropriate anticoagulant use. In elderly patients, drugs associated with medication errors often affect the central nervous system and required dosage adjustments are often neglected [60].

In ambulatory care, specific drug formulations with complex handling requirements promote drug administration errors. For instance, on average, about one in three patients incorrectly self-administers the inhalation device for chronic asthma treatment [61].

Examples by Setting

In adult *inpatients*, administering the wrong dosage is the most frequent medication error. Patients with multiple co-morbidities may

require a dosage adjustment. In pediatric inpatients, wrong dosage often results from dose calculation errors, including 10-fold errors [62]. Moreover, less severe medication errors often result from incomplete drug orders (i.e., not specifying the route of administration if only one route is applicable). However, especially in developed countries, most potential medication errors are intercepted by hospital pharmacists while processing the order.

In the outpatient setting, many medication errors happen at the stage of drug monitoring (e.g., neglecting a required check-up of laboratory values) because patients tend to see their physicians only irregularly. Moreover, they will generally see several physicians concurrently who most often are only partially aware of the actions of their colleagues. Among elderly patients treated in the outpatient setting, the number of physicians seen by a patient was found to be an independent risk factor for a medication error-related ADE [63]. Because patients might often receive drugs from several physicians and additionally purchase over-the-counter drugs, the documentation of an actual and complete medication list is challenging to maintain. Thus, prescription of interacting drugs is frequent and drug–drug interactions contribute to 6% of ADE-related hospital admissions [31]. Compared to the inpatient setting, in the outpatient setting, prescription errors are less likely to be intercepted, so the patient must play a more active role in their medical treatment and assume some degree of responsibility for appropriate drug administration. Two major factors might impede appropriate drug administration: (1) patient nonadherence to prescribed drugs (see also Chapter 38), and (2) inadequate patient knowledge regarding administration, increasing the likelihood of administration errors (e.g., for asthma inhalers).

Moreover, due in part to the fact that information on drug prescription, dispensing, and administration may not be linked, dispensing errors are also important. In a large outpatient study, incidence rates were reported to be about four errors per 10 000 items dispensed [64]. Additionally, inappropriate splitting of tablets was found to be the source of some medication errors [65].

In the *ICU*, critically ill patients are characterized by rapidly changing clinical conditions, receive close and intensive patient monitoring, and require rapid adaptations of their drug therapies. Due to the large number of necessary medications, the frequency of DDIs is particularly high, with about two-thirds of patients having at least one DDI and 44% suffering from a DDI-related ADR in one study [66]. Moreover, a substantial fraction of drugs is given intravenously (IV), potentially using identical IV lines. In one study including 50 ICU patients, 5.8% of concurrently given IV medications were incompatible [67].

In the *long-term care* setting, relatively few data are available [68]. However, medication errors appear to be concentrated in a few different drug classes, most often involving drugs affecting the central nervous system or analgesics [59]. Pharmacotherapy in the elderly occurs in a patient population that is in general multimorbid, polymedicated, and with physiological changes requiring complex dosage adjustment. Therefore, prescribing errors involving inappropriate drug choice as well as inappropriate dosages are frequent [69].

In conclusion, a multitude of different combination of risk factors is possible and these factors must be carefully considered in designing and analyzing pharmacoepidemiology research on medication errors.

Currently Available Solutions

Developing Prevention Strategies

Medication error prevention strategies may address the persons involved in the medication process, the products used, or the process

organization itself. Most often, a prevention strategy might also cover several or all categories; for example, a workflow change will typically include education and training of the staff. Obviously, the best prevention strategies for medication errors will depend on the setting and the nature of the medication errors involved. Slips and lapses in executing correct planned actions can be addressed by workflow changes including skill training and monitoring (e.g., coworker confirmation, checklists) [70]. In contrast, mistakes might be prevented by providing relevant knowledge at the time it is required. Approaches might include educational training as well as provision of paper- or computer-based information at the point of care.

With the majority of errors being knowledge based and occurring during drug prescribing, the implementation of electronic prescribing systems (CPOE) with integrated clinical decision support systems (CDSS) assumes a key role in medication error prevention [71]. Implementation of such systems might eliminate several types of errors, such as transcribing errors [72], and reduce others. Their impact on ADEs in research studies has been less pronounced [73], partly due to the fact that most studies using this approach have been underpowered. However, a metaanalysis from 2014 suggested that implementing CPOE is associated with a greater than 50% decline in the preventable ADE rate [74]. Nearly all CPOE applications in use now were commercially developed, while many of the early studies were done on internally developed systems. In one study, commercial applications in the ICU setting were found in a metaanalysis to be associated with an 85% reduction in the prescribing medication error rates, and a 12% reduction in ICU mortality [75]. Another vulnerable area is intravenous admixture [76]; this is another place where technology is likely to help in the future.

Electronic solutions have been developed to safeguard against drug dispensing or administration errors. For example, barcoding systems are currently used to prevent medication administration to the wrong patient. Electronic medication administration records can be used to electronically monitor drug administration and effectively reduce errors of omission [77]. In the US, CPOE data linked with decision support and barcoding data have become the norm in over 90% of hospitals.

Outcome Assessment

The outcome of prevention strategies is often reported as changes in the frequency of medication errors. However, such information will imperfectly apply as a predictor for health-related outcome. Indeed, in studies assessing both medication errors and patient outcomes, a reduction in medication errors would not necessarily be accompanied by an improvement in patient outcomes. For example, a computer-assisted disease management system might enhance the number of guideline-conformed screenings but the disease severity would not be ameliorated [78]. The assessment of patient outcomes, either by measuring surrogate endpoints (e.g., lab values, disease monitoring parameters) or by assessment of clinical endpoints (e.g., ADE rates, mortality rates), is therefore preferable to estimate the impact of a prevention strategy.

Evaluation of Intervention Strategies

Most prevention strategies are evaluated in a before vs after implementation setting and only scarcely evaluated in randomized trials. Therefore, neglecting of confounding variables can substantially bias the results. In 2005, Han *et al.* reported that the implementation of a CPOE system was an independent factor associated with increasing mortality rates of pediatric inpatients (odds ratio 3.28; 95% confidence interval 1.94–5.55) [79]. However, in this study, the analysis did not control for workflow or policy changes that coincided with the

implementation of the CPOE. Nevertheless, the implementation of prevention strategies might potentially be associated with the introduction of new, "e-iatrogenic" errors [80,81], due to potential changes in workflows. Implementation of CPOE should therefore follow a stepwise roll-out after careful testing and be accompanied by close monitoring [82].

The Future

In the past decades, a multitude of small and several large-scale studies have been conducted in order to assess the frequency and nature of medication errors as well as to evaluate the impact of different prevention strategies. While all studies have found that medication errors happen with considerable frequency during drug therapy, variation in detection approaches makes it hard to narrowly define their incidence and severity. The frequency, best detection approaches, and prevention methods vary by setting and patient population. To allow comparison among study results, careful consideration of the study methodology is especially important. Especially in large-scale studies using only administrative data, information relevant to reliably identifying medication errors is often not systematically available. Key factors in conducting valid research related to medication errors include the consistent use of definitions and classifications of medication errors, and attempts to merge large medication databases with electronic data on patient's clinical information. Another area is refining the decision support in commercial applications.

But perhaps the major current research gap is to develop better approaches for and studies of detection and prevention in the ambulatory care setting – the setting in which the main part of drug treatment takes place. However, additional research is needed in all settings, especially in special populations such as psychiatry and pediatrics.

References

1 Cypress BW. *Drug Utilization in Office Visits to Primary Care Physicians: National Ambulatory Medical Care Survey, 1980*. Department of Health and Human Services publication (PHS) 82–1250. Washington, DC: Public Health Service, 1982.

2 Bates DW, Boyle DL, Vander Vliet MB, Schneider J, Leape LL. Relationship between medication errors and adverse drug events. *J Gen Intern Med* 1995; **10**: 199–205.

3 Institute of Medicine, Kohn LT, Corrigan JM, and Donaldson MS. *To Err is Human. Building a Safer Health System*. Washington, DC: National Academies Press, 1999.

4 Starfield B. Is US health really the best in the world. *JAMA* 2000; **284**: 483–5.

5 Bates DW, Cullen D, Laird N, *et al*. Incidence of adverse drug events and potential adverse drug events: implications for prevention. *JAMA* 1995; **274**: 29–34.

6 Aronson JK. Medication errors: definitions and classification. *Br J Clin Pharmacol* 2009; **67**: 599–604.

7 Krähenbühl-Melcher A, Schlienger R, Lampert M, Haschke M, Drewe J, Kraehenbuehl S. Drug-related problems in hospitals: a review of the recent literature. *Drug Saf* 2007; **30**: 379–407.

8 Leape LL. Preventing adverse drug events. *Am J Health Syst Pharm* 1995; **52**: 379–82.

9 Nebeker J, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. *Ann Intern Med* 2004; **140**: 795–801.

10 Montesi G, Lechi A. Prevention of medication errors: detection and audit. *Br J Clin Pharmacol* 2009; **67**: 651–5.

11 Carpenter JD, Gorman PN. Using medication list – problem list mismatches as markers of potential errors. *AMIA Ann Symp Proc* 2002; **106**.

12 Seidling HM, Schmitt SP, Bruckner T, *et al*. Patient-specific electronic decision support reduces prescription of excessive doses. *Qual Saf Health Care* 2010; **19**(5): e15.

13 Gandhi TK, Weingart SN, Borus J, *et al*. Adverse drug events in ambulatory care. *N Engl J Med* 2003; **348**: 1556–64.

14 Vrbnjak D, Denieffe S, O'Gorman C, Pajnkihar M. Barriers to reporting medication errors and near misses among nurses: a systematic review. *Int J Nurs Stud* 2016; **63**: 162–78.

15 Potylycki MJ, Kimmel SR, Ritter M, *et al*. Nonputitive medication error reporting: 3-year findings from one hospital's primum non nocere initiative. *J Nurs Adm* 2006; **36**: 370–6.

16 Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. *Jt Comm J Qual Improv* 1995; **21**: 541–8.

17 Kopp BJ, Erstad BL, Allen AM, Theodorou AA, Priestley G. Medication errors and adverse drug events in an intensive care unit: direct observation approach for detection. *Crit Care Med* 2006; **34**: 415–25.

18 Dean B, Barber N. Validity and reliability of observational methods for studying medication administration errors. *Am J Health Syst Pharm* 2001; **58**: 54–9.

19 Metzger J, Welebob E, Bates DW, Lipsitz S, Classen DC. Mixed results in the safety performance of computerized physician order entry. *Health Affairs* 2010; **29**: 655–63.

20 Rieger K, Scholer A, Arnet I, *et al*. High prevalence of unknown co-medication in hospitalised patients. *Eur J Clin Pharmacol* 2004; **60**: 363–8.

21 Kuperman GJ, Gandhi TK, Bates DW. Effective drug-allergy checking: methodological and operational issues. *J Biomed Inform* 2003; **36**: 70–9.

22 Benkhaial A, Kaltschmidt J, Weisshaar E, Diepgen TL, Haefeli WE. Prescribing errors in patients with documented drug allergies: comparison of ICD-10 coding and written patient notes. *Pharm World Sci* 2009; **31**: 464–72.

23 Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *JAMA* 1997; **277**: 301–6.

24 Hug BL, Witkowski DJ, Sox CM, *et al*. Adverse drug event rates in six community hospitals and the potential impact of computerized physician order entry for prevention. *J Gen Intern Med* 2010; **25**: 31–8.

25 Kane-Gill SL, Kowiatek JG, Weber RJ. A comparison of voluntary reported medication errors in intensive care and general care units. *Qual Saf Health Care* 2010; **19**: 55–9.

26 Benkirane RR, Abouqal R, Haimeur CC, *et al*. Incidence of adverse drug events and medication errors in intensive care units: a prospective multicenter study. *J Patient Saf* 2009; **5**: 16–22.

27 Greene SB, Williams CE, Pierson S, Hansen RA, Carey TS. Medication error reporting in nursing homes: identifying targets for patient safety improvement. *Qual Saf Health Care* 2010; **19**: 218–22.

28 Ferrah N, Lovell JJ, Ibrahim JE. Systematic review of the prevalence of medication errors resulting in hospitalization and death of nursing home residents. *J Am Geriatr Soc* 2017; **65**: 433–42.

29 Mira JJ, Lorenzo S, Guilabert M, Navarro I, Pérez-Jover V. A systematic review of patient medication error on self-administering medication at home. *Expert Opin Drug Saf* 2015; **14**: 815–38.

30 Parand A, Garfield S, Vincent C, Franklin BD. Carers' medication administration errors in the domiciliary setting: a systematic review. *PLoS One* 2016; **11**: e0167204.

31 McDonnell PJ, Jacobs MR. Hospital admissions resulting from preventable adverse drug reactions. *Ann Pharmacother* 2002; **36**: 1331–6.

32 Gandhi TK, Weingart SN, Seger AC, *et al*. Outpatient prescribing errors and the impact of computerized prescribing. *J Gen Intern Med* 2005; **20**: 837–41.

33 Medonca JM, Lyra DP Jr, Rabelo JS, *et al*. Analysis and detection of dental prescribing errors at primary health care units in Brazil. *Pharm World Sci* 2010; **32**: 30–5.

34 Shiu JR, Fradette M, Padwal RS, *et al*. Medication discrepancies associated with a medication reconciliation program and clinical outcomes after hospital discharge. *Pharmacotherapy* 2016; **36**: 415–21.

35 Quélennec B, Beretz L, Paya D, *et al*. Potential clinical impact of medication discrepancies at hospital admission. *Eur J Intern Med* 2013; **24**: 530–5.

36 Shah K, Barker KA. Out-of-hospital medication errors: a 6-year analysis of the national poison data system. *Pharmacoepidemiol Drug Saf* 2009; **18**: 1080–5.

37 Mesteig M, Helbostad JL, Sletvold O, Røsstad T, Saltveldt I. Unwanted incidents during transition of geriatric patients from hospital to home: a prospective observational study. *BMC Health Serv Res* 2010; **10**: 1.

38 Miller MR, Robinson KA, Lubomski LH, Rinke ML, Pronovost PJ. Medication errors in paediatric care: a systematic review of epidemiology and an evaluation of evidence supporting reduction strategy recommendations. *Qual Saf Health Care* 2007; **16**: 116–26.

39 Chedoe I, Molendijk HA, Dittrich ST, *et al*. Incidence and nature of medication errors in neonatal intensive care with strategies to improve safety: a review of the current literature. *Drug Saf* 2007; **30**: 503–13.

40 Walsh KE, Dodd KS, Seetharaman K, *et al*. Medication errors among adults and children with cancer in the outpatient setting. *J Cin Oncol* 2009; **27**: 891–6.

41 Hanlon JT, Schmader KE, Koronkowski MJ, *et al*. Adverse drug events in high risk older outpatients. *J Am Geriatr Soc* 1997; **45**: 945–8.

42 Onder G, Pedone C, Landi F, *et al*. Adverse drug reactions as a cause of hospital admissions: results from the Italian Group of Pharmacoepidemiology in the Elderly (GIFA). *J Am Geriatr Soc* 2002; **50**: 1962–8.

43 Fick DM, Cooper JW, Wade WE, Waller JL, Maclean JR, Beers MH. Updating the Beers criteria for potentially inappropriate medication use in older adults: results of a US consensus panel of experts. *Arch Intern Med* 2003; **163**: 2716–24.

44 Steinman MA, Rosenthal GE, Landefeld S, Bertenthal D, Kaboli PJ. Agreement between drugs-to-avoid criteria and expert assessments of problematic prescribing. *Arch Intern Med* 2009; **169**: 1326–32.

45 Flynn EA, Barker KN, Pepper GA, Bates DW, Mikeal RL. Comparison of methods for detecting medication errors in 36 hospitals and skilled-nursing facilities. *Am J Health Syst Pharm* 2002; **59**: 436–46.

46 Kozer E, Scolnik D, Jarvis AD, Koren G. The effect of detection approaches on the reported incidence of tenfold errors. *Drug Saf* 2006; **29**: 169–74.

47 Brown CA, Bailey JH, Lee J, Garrett PK, Rudman WJ. The pharmacist–physician relationship in the detection of ambulatory medication errors. *Am J Med Sci* 2006; **331**: 22–4.

48 Kaushal R, Bates DW, Landrigan C, *et al*. Medication errors and adverse drug events in pediatric inpatients. *JAMA* 2001; **285**: 2114–20.

49 Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. *Ann Intern Med* 2003; **138**: 161–7.

50 Thomsen LA, Winterstein AG, Søndergaard B, Haugbølle LS, Melander A. Systematic review of the incidence and characteristics of preventable adverse drug events in ambulatory care. *Ann Pharmacother* 2007; **41**: 1411–26.

51 Bates DW, Miller EB, Cullen DJ, *et al*. Patient risk factors for adverse drug events in hospitalized patients. ADE Prevention Study Group. *Arch Intern Med* 1999; **159**: 2553–60.

52 Keers RN, Williams SD, Cooke J, Ashcroft DM. Causes of medication administration errors in hospitals: a systematic review of quantitative and qualitative evidence. *Drug Saf* 2013; **36**: 1045–67.

53 Meyer-Massetti C, Meier CR, Guglielmo BJ. The scope of drug-related problems in the home care setting. *Int J Clin Pharm* 2018; **40**(2): 325–34.

54 Winterstein AG, Johns TE, Rosenberg EI, Hatton RC, Gonzalez-Rothi R, Kanjanarat P. Nature and causes of clinically significant medication errors in a tertiary care hospital. *Am J Health Syst Pharm* 2004; **61**: 1908–16.

55 Kanjanarat P, Winterstein AG, Johns TE, Hatton RC, Gonzalez-Rothi R, Segal R. Nature of preventable adverse drug events in hospitals: a literature review. *Am J Health Syst Pharm* 2003; **60**: 1750–9.

56 Hennipman B, de Vries E, Bökkerink JP, Ball LM, Veerman AJ. Intrathecal vincristine: 3 fatal cases and a review of the literature. *J Pediatr Hematol Oncol* 2009; **31**: 816–19.

57 Noble DJ, Donaldson LJ. The quest to eliminate intrathecal vincristine errors: a 40-year journey. *Qual Saf Health Care* 2010; **19**(4): 323–6.

58 Mohr JF, Hall AC, Ericsson CD, Ostrosky-Zeichner L. Fatal amphotericin B overdose due to administration of nonlipid formulation instead of lipid formulation. *Pharmacotherapy* 2005; **25**: 426–8.

59 Winterstein AG, Hatton RC, Gonzalez-Rothi R, Johns TE, Segal R. Identifying clinically significant preventable adverse drug events through a hospital's database of adverse drug reaction reports. *Am J Health Syst Pharm* 2002; **59**: 1742–9.

60 Hansen RA, Grenne SB, Williams CE, *et al*. Types of medication errors in North Carolina nursing homes: a target for quality improvement. *Am J Geriatr Pharmacother* 2006; **4**: 52–61.

61 Hagmolen of ten Have W, van de Berg NJ, Bindels PJ, van Aalderen WM, van der Palen J. Assessment of inhalation technique in children in general practice: increased risk of incorrect performance with new device. *J Asthma* 2008; **45**: 67–71.

62 Lesar TS. Tenfold medication dose prescribing errors. *Ann Pharmacother* 2002; **36**: 1833–9.

63 Green JL, Hawley JN, Rask KJ. Is the number of prescribing physicians an independent risk factor for adverse drug events in an elderly outpatient population? *Am J Geriatr Pharmacother* 2007; **5**: 31–9.

64 Ashcroft DM, Quinlan P, Blenkinsopp A. Prospective study of the incidence, nature and causes of dispensing errors in community pharmacies. *Pharmacoepidemiol Drug Saf* 2005; **14**: 327–32.

65 Quinzler R, Szecsenyi J, Haefeli WE. Tablet splitting: patients and physicians need better support. *Eur J Clin Pharmacol* 2007; **63**: 1203–4.

66 Bertsche T, Pfaff J, Schiller P, *et al*. Prevention of adverse drug reactions in intensive care patients by personal intervention based on an electronic clinical decision support system. *Intensive Care Med* 2010; **36**: 665–72.

67 Bertsche T, Mayer Y, Stahl R, Hoppe-Tichy T, Encke J, Haefeli WE. Prevention of intravenous drug incompatibilities in an intensive care unit. *Am J Health Syst Pharm* 2008; **65**: 1834–40.

68 Barker KN, Flynn EA, Pepper GA, Bates DW, Mikeal RL. Medication errors observed in 36 health care facilities. *Arch Intern Med* 2002; **162**: 1897–903.

69 Fialova D, Onder G. Medication error in elderly people: contributing factors and future perspectives. *Br J Clin Pharmacol* 2009; **67**: 641–5.

70 Conroy S, Sweis D, Planner C, *et al*. Intervention to reduce dosing errors in children: a systematic review of the literature. *Drug Saf* 2007; **30**: 1111–25.

71 Anton C, Nightingale PG, Adu D, Lipkin G, Ferner RE. Improving prescribing using a rule based prescribing system. *Qual Saf Health Care* 2004; **13**(3): 186–90.

72 Bates DW, Leape LL, Cullen DJ, *et al*. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *J Am Med Inform Assoc* 1998; **280**: 1311–16.

73 Garg AX, Adhikari NK, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005; **293**: 1223–38.

74 Nuckols TK, Smith-Spangler C, Morton SC, *et al*. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. *Syst Rev* 2014; **3**: 56.

75 Prgomet M, Niazkhani Z, Georgiou A, Westbrook JI. Impact of commercial computerized provider order entry (CPOE) and clinical decision support systems (CDSSs) on medication errors, length of stay, and mortality in intensive care units: a systematic review and meta-analysis. *J Am Med Inform Assoc* 2017; **24**: 413–22.

76 Hedlund N, Beer I, Hoppe-Tichy T, Trbovich P. Systematic evidence review of rates and burden of harm of intravenous admixture drug preparation errors in healthcare settings. *BMJ Open* 2017; **7**: e015912.

77 Scott-Cawiezell J, Madsen RW, Pepper GA, Vogelsmeier A, Petroski G, Zellmer D. Medication safety teams' guided implementation of electronic medication administration records in five nursing homes. *Jt Comm J Qual Patient Saf* 2009; **35**: 29–35.

78 Meigs JB, Cagliero E, Dubey A, *et al*. A controlled trial of Web-based diabetes disease management: the MGH Diabetes Primary Care Improvement Project. *Diabetes Care* 2003; **26**: 750–7.

79 Han YY, Carcillo JA, Venkataraman ST, *et al*. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005; **116**: 1506–12.

80 Weiner JP, Kfuri T, Chan K, Fowles JB. "e-iatrogenesis": the most critical unintended consequence of CPOE and other HIT. *J Am Med Inform Assoc* 2007; **14**: 387–8.

81 Koppel R, Metlay JP, Cohen A, *et al*. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005; **293**: 1197–203.

82 Rosenbloom ST, Harrell FE Jr, Lehmann CU, Schneider JH, Spooner SA, Johnson KB. Perceived increase in mortality after process and policy changes implemented with computerized physician order entry. *Pediatrics* 2006; **117**: 1452–5.

42

Patient Engagement and Patient-Reported Outcomes

Esi M. Morgan

Department of Pediatrics, Division of Rheumatology, Cincinnati Children's Hospital Medical Center, and University of Cincinnati College of Medicine, Cincinnati, OH, USA

Patient engagement reflects the effective participation of patients in their own healthcare. A historical view of the relationship of patient to clinician (physician, provider, etc.), may have positioned the patient in the role of a provider of medical history and symptoms, and subsequently a recipient of a "healthcare product," the product being clinician-generated information, advice, treatment prescription, and instructions that the patient was expected to follow. The outcome of care was thus a factor of the quality of the advice and treatment recommendations as well as the ability and inclination of the patient to carry out this advice.

The influential 2001 report by the Institute of Medicine on reforming and improving healthcare in the US named patient-centeredness as one of six guiding aims for healthcare, which should also be safe, effective, timely, efficient, and equitable [1]. It has been argued that healthcare is better conceived of as a service rather than a product provided to patients, and as such achievement of best patient outcomes may hinge on taking a co-production approach to care. In this service-oriented (rather than transactional) view, the patient (service user) and clinician (service professional) are in a collaborative relationship with shared creation of the healthcare service and shared responsibility for outcomes. Effectively, the healthcare service is "co-produced" [2]. Self-management support is a key component of effective chronic illness care management that is patient centered [3]. Concepts such as effective communication, shared decision making and monitoring of health outcomes to support successful self-management become important and essential goals for co-production of patient-centered care, particularly for chronic conditions (see Chapter 39).

The reframing of the patient–clinician relationship as co-producers of a healthcare service elevates the importance of measuring PROs in healthcare delivery. By extension, it increases the utility of PRO use in observational studies and clinical trials in order for patients to be able to translate expected benefit of study medications to expected experience in real-world use based on meaningful clinical outcome assessments. Broader use of standardized PRO measures in healthcare also increases the feasibility of conducting comparative effectiveness research with observational clinical data.

Patient engagement in comparative effectiveness research more generally, with patients moving from the role of study subject to becoming partners in prioritizing research questions

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

and contributing to study design, has become a phenomenon over the past few decades due to patient advocacy efforts. The Patient Centered Outcomes Research Institute (PCORI), which has been funding research since 2012, has promoted engagement of multiple stakeholders (patients, clinicians, payers, healthcare institutions, etc.) in the conduct of clinical research and is systematically studying the impact of patient engagement on the research processes. Patient stakeholders on research teams may influence research topics towards patient priorities and clinical relevance, contribute to study design including selection of outcome measures, help with study materials including consent forms, participate in recruitment and retention efforts, reduce missingness in data collection, and provide important perspectives when interpreting study results. Given the brief period of time that patient-centered approaches to research have been emphasized, it is too early to assess the impact of patient engagement on study completion or uptake of findings into clinical practice [4].

Patient-reported outcomes (PROs) are defined as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [5]. PROs enable patients to directly communicate information about their status with respect to various domains of health, such as symptoms, function, quality of life and participation, as well as health behaviors and experience with care, in a structured format that translates the patient experience into data. Use of PROs recognizes that the health concerns of the patient may not be adequately perceived by the clinician unless there is direct and structured inquiry of what is bothering the patient. As a consequence of failing to elicit direct patient input, clinicians may prioritize other health topics less valued by the patient. Incorporation of PROs into clinical care has the potential to enrich clinician–patient communication as the patient voice is systematically and routinely collected, as well as providing data in the form of meaningful scores to which clinical decisions and interventions can be anchored and that may be complementary to lab tests or physical exam. Building a system around incorporation of PROs into the practice of medicine supports patient engagement.

Advances in measurement theory and the use of item response theory (IRT) models in scale development have enabled development of modern PRO measures that are both more efficient (shorter) and have more precision in scoring [6] than questionnaires developed with classic test theory. In support of the clinical research enterprise, the National Institutes of Health supported the development and distribution of various IRT-based PROs, including the Patient Reported Outcomes Measurement Information System (PROMIS®) which started as a cooperative network in 2004 to develop unidimensional measures of physical, mental, and social health domains available as short forms, item banks, and computer adaptive tests. PROMIS measures were designed to be generic and used across chronic conditions to support clinical research [7], and the use case evolved to support outcomes assessment in clinical care. Numerous examples of other PROs have been widely used in research studies, and increasingly in clinical care.

Conceptualization of PROs as an essential component to patient engagement in healthcare has the power not only to influence and improve outcomes, but potentially to transform the relationship of patients to the perception of their own health and agency, and their communication and relationship with their healthcare professionals [8]. The vision is compelling but the reality of implementation, interpretation, use, and evaluation poses practical and methodologic challenges. This chapter will explore these challenges in the context of integration of PROs into pharmacoepidemiology research, including clinical trials and observational studies.

Patient-Reported Outcomes in Clinical Trials

It is increasingly recognized that including the perspective of the patient end-user is important during development of new medication or medical products. Capturing the patient perspective as a clinical trial outcome may take the form of a PRO. PROs can be rating scales, symptoms reports, or questionnaires completed by self-report or completed in the form of a structured or unstructured interview, provided there is no modification of the patient's response by the interviewer. Whereas only PROs can be used to capture symptoms and other unobservable domains of health, PROs can also be used to capture and include the patient viewpoint on observable domains (e.g., functional ability, counts of events). Methodologic challenges in longitudinal measurement and interpretation of PROs are discussed later.

In addition to PROs, there are three other forms of clinical outcome assessments for use in trials: clinician-reported outcomes, observer-reported outcomes, and performance outcomes. *Clinician-reported outcomes* are measurements completed by healthcare professionals based on their evaluation of the patient's health condition. By definition, these are observable and include physical findings and report of clinical events or interpretation of clinical data supporting occurrence of such events (lab or other physiologic tests). *Observer-reported outcomes* are also based on observable signs, events or behaviors related to the patient's health condition; however, they do not require interpretation or report by a healthcare professional but rather, observation can be by a caregiver, such as a parent. Examples could include event counts or behavior rating scales. Last, *performance outcomes* are based on patient completion of tasks as directed by a healthcare professional, for example, timed walking distance (gait speed) or tests of cognition/memory (word recall) [9]. These measurements are valued in part when they are understood by medical decision makers, clinicians and patients to be clinically meaningful endpoints [10].

Patient-reported outcomes serve a variety of functions in new product development. They can help better understand the impact of a medical condition on the patient, such as in the realm of physical, mental, or social health, and also uncover if there are (unmet) medical needs to be addressed by a new medication [11]. PROs can also be used to enable estimation of the benefit to the patient of use of a medication during the clinical trial (e.g., by estimating the minimal clinically important treatment effect or minimal clinically important difference [MCID]), and by extension, can be used to characterize the expected benefit to the end-user in the clinical setting. Adverse effects of treatment, or an estimate of risks of use, may also be captured from the patient perspective [5,10]. The Food and Drug Administration (FDA) as part of the 2012 reauthorization of the Prescription Drug User Fee Act (PDUFA V) has been including patient input on the impact of chronic illness to inform which aspects of illness would be most meaningful to target from the patient perspective. Ascertaining such information from cohorts of patients with chronic conditions and compiling perspectives on meaningful clinical impact that is desired from new products require a systematic approach and robust methods to interpret and use the information to inform drug development [11].

After clinical outcomes that are meaningful drug targets from the patient perspective are identified, it becomes important to be able to determine whether the drug/treatment/intervention under investigation exerts a clinically meaningful effect on the relevant clinical outcome. Although a statistically significant change in outcome measure scores may be achieved, achieving statistical significance may not correspond to a clinically meaningful benefit. Determination that a change is of clinical meaningfulness rests on the idea that the outcome being measured is of relevance and importance

to the patient to begin with, and that the degree of change exerted creates a clinically appreciable benefit (reduction in symptoms, increase in function, greater wellbeing, etc.). Understanding how to estimate clinically meaningful change in PROs is an area of active methodological research [12,13].

Patient-Reported Outcomes in Routine Care

A current challenge to use of PROs as outcomes in observational studies is that such measures are not typically available, especially when compared to measuring other outcomes such as clinical events or death. However, PRO measures are growing in use in clinical care in part due to the push for patient engagement to provide patient-centered care and improve healthcare value [14,15]. Conceptually, capturing PROs in clinical care could be seen as a method to include the patient voice in outcome estimation. In practice, incorporation of PROs is starting to effect meaningful healthcare improvement. Over the past decade, there has been a growing body of published evidence on the benefits of incorporating PROs in clinical care, including both process and outcomes improvement. A number of technical solutions for full integration into the electronic health record, used alongside the electronic health record or hybrid approaches, have been described [16]. User's guides have been published with step-by-step instructions for health systems considering adoption of PROs to facilitate incorporation into clinical care [16,17].

Better collection and understanding of PROs can also lead to interventions to improve outcomes. Published reports of integration of PROs into clinical care have over time evolved from description of acceptability and feasibility [18], to demonstration of improved care processes [19] and communication, to preliminary evidence of increased clinician satisfaction and even improved outcomes. When PROs are collected and reviewed, patients perceive the clinician to have increased awareness of their symptoms, which otherwise might go unrecognized and unaddressed [20]. Clinicians may be better able to identify and diagnose patient conditions, including mental health concerns, if PRO results are routinely collected [8]. Use of PROs may result in increased clinician awareness of the impact of the health condition on the patients' health-related quality of life and facilitate patient–clinician discussions [21]. Communication [22], shared decision making, and collaborative treatment planning are enhanced when PROs are incorporated into patient and clinician conversations [15,18]. Until recently, despite theoretical support for the incorporation of PROs in clinical practice to improve outcomes, there has been scant evidence of demonstrable benefit to outcomes. As PROs become more established and reliably implemented in routine clinical practice, researchers are moving beyond studies of process improvement to studies of the impact of PROs on outcomes of care.

Much of the literature on the effective use of PROs in clinical medicine has come from the fields of oncology and surgery. A recent compelling report from an oncology randomized trial of PRO use during routine cancer treatment showed that integration of PROs in the form of electronic symptom monitoring into care of patients with metatastic cancer was associated with increased survival [23]. The authors postulate that this could be a consequence of expedited medical care team response to patient symptoms, which could result in two potential benefits. First, averting potential adverse events with clinical interventions including adjustments to chemotherapy dose, and second, administering chemotherapy for a relatively longer duration which was possible due to reduction in adverse symptoms [24] prompted by early ascertainment and response to the symptoms. This brief research letter [23] illustrates a compelling account of the potential

benefit of PROs in improved management of patient symptoms and survival; however, the impact on alleviating adverse symptoms and reduction of suffering is meaningful in itself. Bringing increased survival into the calculus enhances the urgency of integration of PRO assessment – review of scores and taking appropriate action – into delivery of healthcare.

Recently, payers have been encouraging PRO collection [25], including the use of financial incentives for tracking PRO data. For example, PROs are used to evaluate the impact and trajectory for improvement of surgical interventions in the case of Medicare reimbursement for elective joint replacement [26]. The availability of such data can enhance observational studies by providing prospectively collected information on PROs.

Patient-Reported Outcomes as Motivation to Develop New Therapeutic Strategies

From a therapeutic perspective, incorporation of PROs may contribute to a more targeted conversation on issues of concern for the patient that might otherwise go unaddressed. It may help to quantify the discomfort a patient feels and subsequently trigger the use of a therapeutic intervention. Information on disease control status, such as number of asthmatic episodes, can focus the visit on topics of most relevance and be a useful measure of disease burden. Having prespecified and agreed-upon PRO score thresholds that trigger specific evidence-based interventions facilitates action by the clinician. Alternatively, aggregate patient-reported data on symptoms, function, quality of life, or experience related to an intervention can be shared with an individual patient as part of a shared decision-making discussion [26]. Several examples are provided here.

Self-management support is a key component of the Chronic Care Model, a framework that has been useful in driving health system improvement [3]. Patient engagement, with informed, activated patients, is key to productive patient–clinician relationships which effect clinical outcome improvement. Successful self-management support programs have the following elements:

- clinicians communicate the expectation for the central role of the patient in managing their condition
- patients' self-management skills, confidence in management ability, barriers and supports are assessed regularly
- trained staff employ behavior change interventions
- patients co-develop with healthcare professionals their own individualized treatment plans and support is available on an ongoing basis if the patient needs it [27].

Self-management support is a patient-centered, iterative, and ongoing process. Experience from a quality improvement collaborative on implementing self-management support revealed that the most successful care teams received relevant training, adopted a philosophy of patient-centered care and collaborative goal setting, integrated components into the care delivery system, and assigned accountable staff [27]. Due to the fact that the majority of chronic illness care occurs outside the medical office in the interval (days to months) between office visits, PRO assessment has the potential to become an effective facilitator of between-visit communication between patients and clinicians and may allow patients to better manage their health conditions.

Since medical care at outpatient office visits occurs at relatively infrequent intervals while health events occur on an ongoing basis, the trajectory of illness may not be adequately captured at clinic visits due to factors such as – on the patient side – the failings of memory (recall), the lack of a way to accrue and organize data, and perhaps – on the clinician side – failure to inquire (particularly in the case where PROs are not assessed). As a result, medical decision

making may not incorporate all relevant health information into treatment decisions. Imagine the situation where Patient A begins a new medication, Drug A, and experiences rapid improvement in signs and symptoms of disease which gradually return to baseline over time. Patient B experiences gradual improvement following the initiation of Drug B, then rapid worsening before the condition returns to baseline by the time of the follow-up visit. Both patients have roughly the same disease activity level at the follow-up visit. In this scenario, information is lost about the relative lack of effectiveness and negative impact Patient B experienced being on Drug B, as more time was spent in a flare state than with Patient A who took Drug A.

Technology can enable the capture and transmission of both patients' unique experience to the clinical team. Data can be collected at home, either as PRO questionnaires on computer or mobile devices, or as technology-enabled patient-generated data (e.g., physiologic monitors), and then shared with the clinical team. Data collection from wearable devices requires little or no effort from the patient perspective. This opens the possibility for patient data from in-between patient visits to serve action-oriented interventions. However, numerous potential barriers exist to such a system: availability of optimal technology platform to house and transmit data, cost of technology, and patient having access to a reliable and persistent Wi-Fi connection. From the clinician perspective, additional considerations include the interoperability of the technology with the electronic health record, providing training to use the system and interpret results, allocating staff to review the results and integrating this process into the clinical workflow. Patients who transmit data will have the expectation that results are reviewed, so workflows will need to be optimized to allow for reliability of review, and prompt response if deterioration in status or other change is identified.

Such a system is currently being tested in pediatric chronic illness care, with, for example, the Orchestra mobile health technology platform and care model intervention in cystic fibrosis and inflammatory bowel disease [28]. The concept, currently being piloted, is of a mobile health (mHealth) technology platform that enables clinicians and patients to work collaboratively together to co-produce healthcare. It does so in part by allowing symptom tracking with real-time data visualization on a platform that is shared with patients and clinicians, allowing for continuation of patient–clinician interaction outside the office visit. Furthermore, the system contains automated symptom surveillance detection software which generates alerts when potentially important changes in health status are detected. Other features supporting patient engagement include previsit planning support and opportunities to participate in research. For example, patients may be recruited to participate in a study of the impact of new medications, other interventions such as lifestyle changes on PROs, or other measures being tracked [28].

There are other insights to be gained from the operationalization of the mHealth technology, including study of the optimal way to integrate such a system into clinical workflow and evaluate the system in terms of feasibility, acceptance, and completion rates. Considering a culture shift is necessary to embrace the co-production model, internal (clinical team) and external (patient) stakeholder engagement and training are required – as are social contracts. Ultimately, the success of the system may be measured in terms of increased patient engagement and self-efficacy (i.e., belief in one's ability to perform activities necessary to achieve health goals), improved outcomes, and appropriate healthcare utilization [28]. The concept is potentially transformative in the ability of patients to self-manage care, including supporting development of individualized treatment plans and customized treatment trials, N-of-1 studies, with care being optimized based on a patient's own PRO data.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Patient-reported outcomes are directly relevant to pharmacoepidemiology research insofar as they capture information central to the estimation of benefit and potential adverse effects of medication that may elude physical exam or lab testing, such as symptoms like fatigue, cognitive impairment, or pain. Clinical trials have restricted sample size and duration of use compared to real-world consumption of medications. Routine and systematic collection of PROs as part of clinical care can help us better understand the epidemiology of disease and, importantly, also the impact of medications on symptoms and quality of life as part of postmarket surveillance. The value of the information is contingent on the reliability of data collection from a representative population, and ability to access the data for research. While PRO collection has benefit for pharmacoepidemiology research, the rationale behind incorporation of the patient voice into clinical care via the use of PROs is to identify unmet needs of patients and help identify gaps in healthcare. Once such a need is identified, ideally an action or response would occur, such as change in medication, referral for treatment, or development of an individualized action plan for lifestyle or behavior change as part of a self-management support. Motivating the move from recognition and identification of the problem to action is a major area for attention.

Ensuring PRO Completion and Results Review

In order to generate valid estimates of treatment effectiveness or impact on quality of life from analysis of PROs in clinical observational data, it is essential that clinical processes are in place to secure research quality data. As the primary motivation for PROs collection may be delivery of optimal patient-centered care and better health outcomes, with the ability to use the data for research as secondary, the clinical team – and patients – must be aligned in perceiving the value of quality PRO data collection. To be most useful clinically, and to allow valid inferences in research, PROs must be collected routinely (reliably at regular intervals), uniformly (in a consistent manner), completely (able to be scored), and correctly (patients understand the questions and answers reflect their health status).

Once a decision is made to incorporate PROs into clinical care, there must be a means to achieve reliable review of the results by the clinical team. This review may not always occur, as reflected in an example from a leading institution in which a discordance was noted between symptom documentation by self-report for signs of heart disease and lack of clinician documentation [29]. If clinicians do not understand the PRO measure, the way it is scored, do not feel it relevant to their specialty, or feel there is no intervention or action that can impact the PRO, they may not be inclined to discuss it [30]. The solution may be cultural and social, and require a clinical champion to motivate colleagues and gain agreement on selection of PROs (more on this later). Or it may require a solution similar to the idea of setting a social contract, noted in the Orchestra co-production example [28]. Barriers may be technical, related to location and ease of viewing of the reports in the medical record, in which case this may require an information technology solution. Another solution may leverage quality improvement methods and hinge on building a workflow in which discussion of PRO completion is an expected part of the clinic visit and is included into process of care quality measures.

There are significant consequences of clinicians not reviewing PROs completed by patients. When clinicians do not review PROs, it becomes a threat to co-production of care,

risks the patient feeling their time and input were not valued, and may result in a decreased willingness for PRO completion at subsequent visits; it is also a missed opportunity to check the reliability of the system for PRO administration (i.e., whether the patient even received the PRO to complete from the staff). Nonreview of PRO results is therefore a threat to patient satisfaction with care, their perceived trust in PROs, and a missed opportunity to improve functional outcomes.

Review of PRO results enables verification of patient understanding of the instrument and that it reflects their health status, increases recognition of areas for health intervention, and provides positive feedback that encourages completion of PROs at subsequent visits.

PRO Selection, Score Interpretation and Interventions

As noted earlier, instrument selection is vitally important for both clinical and research applications to ensure the measure is of clinical relevance, addresses an outcome of interest, and is brief and practical to administer. These and other considerations such as ease of understanding, ability to foster patient–clinician communication, and value in identifying unmet needs should be reviewed with representative patients and clinicians for input and agreement on importance. In addition, the basic prerequisites of adequate psychometric properties such as reliability, validity, and responsiveness should be met [17,25]. A comparison study of various PRO measures in oncology patients with feedback from patients and clinicians helped to highlight the importance of reviewing and selecting PROs specific and useful for the intended application in clinical practice [31].

In order to be actionable, scores on measures for a specific health condition should be known, including the normal range versus when to intervene for an abnormal score. There is need for consensus and standard-setting processes on threshold scores to trigger interventions. In order to track longitudinal change, it is helpful if the minimal clinically important difference for improvement or worsening has been determined (discussed further in methodologic considerations later). Ensuring these scores and thresholds for action are known, and identifying evidence-based interventions or, where such evidence does not exist, gaining consensus agreement amongst practice clinicians and patient stakeholders on recommended interventions will facilitate clinicians and patients moving from simple review and discussion of PROs to taking effective action and co-producing treatment plans based in part on PRO scores which are meaningful to them.

Considerations when selecting PROs for clinical use are likewise important for research applications. Using PROs with known clinically meaningful endpoints will facilitate longitudinal observational studies of effectiveness of medications or interventions.

Patient Engagement and Individualized Assessment and Treatment Plans

Mobile health tracking systems such as described in the Orchestra example [28] offer the possibility of customized measurement (choice of PROs, timing of administration), and use of individualized measures which may be of great importance in pharmacoepidemiology outcome studies [8]. In addition, use of PROs for in-between clinic visit care for continuous monitoring and data feedback, with customized reporting and means to detect signals of health status change in a personalized system, lends itself to N-of-1 trials. In this type of study, an individual can make planned changes to treatment or lifestyle modification, then track changes in symptoms to evaluate for possible effectiveness. By keeping daily journals or other means of annotation, patients can see if other triggers or environmental factors may have led to changes in their health status [28]. This structured type of intervention allows

study of the impact of medications with relatively rapid onset and dissipation of effects (e.g., pain medications). In addition, one can study whether patients who are given access to their outcome data become more engaged in their health such as being more prepared for clinic visits and involved in shared decision making and treatment planning.

Barriers to Measuring PROs in Clinical Practice and Using PROs to Guide Interventions

Barriers to PRO use are varied and start with garnering institutional resources to obtain, store, and update electronic data collection tools. Designing integration of PROs into clinical workflow is a necessary step which first requires gaining clinician consensus on the purpose and value of adding PRO capture and review to the clinical encounter. Clinicians must understand how the use of structured PROs adds value over simply asking the patient how they feel [15]. Frontline staff training on the importance of PRO collection and distribution of questionnaires or collection devices is required for reliable PRO collection.

Clinician training is required on how to use PROs to facilitate high-quality communication with patients. This requires training on how to interpret PRO results and how to communicate about the results with patients. The process could be facilitated by orientating the patients themselves to use of PROs and their role in their clinical care. For PROs to help facilitate co-production of care may in some circumstances necessitate reframing of the patient–clinician relationship, a complex endeavor in culture and behavior change [2].

Buy-in may become easier to obtain as more convincing data become available on PRO integration into care resulting in improved outcomes, as barriers to logistics of PRO collection are lowered, as graphical displays become more intuitive, and as interpretation of data and action steps become more familiar and supported with decision aids [15]. Clinicians may have higher interest in use of PROs if they receive feedback on their own patients' scores, relating to outcomes or experience with care (satisfaction) and recognize they can take steps to improve performance [30].

As use of PROs in clinical trials and pharmacoepidemiology research becomes more prevalent, PRO endpoints may become more widely accepted goals to measure and monitor treatment efficacy. This may serve as positive reinforcement to PRO use for monitoring treatment effectiveness in clinical practice settings.

Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

Just as patient engagement and PROs inform the discussion in the clinical practice setting, they similarly play an increasing role in the research setting. For example, including patient-relevant outcomes in clinical trials has become a priority of the FDA [5,11] and fostering patient engagement in all stages of research has been the genesis of the PCORI and its significant funding portfolio [32]. However, there are methodologic challenges with incorporating PROs into clinical research, including issues of discordance in perception between raters and measurement of within-person change over time.

Discordance in Perspectives Between Patients, Clinicians, and Researchers

With respect to differences in rater perspectives, there are examples of discrepancies between perceptions of clinicians and patients regarding level of disease activity, and about whether there has been improvement or deterioration in the condition. There may also be lack

of concordance between composite measures of disease activity used to assess efficacy in clinical trials and measures that use PROs.

Examples from the field of rheumatology include the comparison of composite indices used in rheumatoid arthritis clinical trials (e.g., complete joint counts, laboratory values, physician global assessment), which tend to be lengthy to complete compared to PROs, which are shorter and more feasible for use in clinical practice. Composite indices are helpful to reduce the presentation of information from multiple measures into a single summary score. This is most informative when the measures included in an index are highly correlated. When a measure does not correlate or track well with others, to include it in a composite index would result in lost or obscured information. While there is a correlation between the results from composite indices and PROs, studies have found that the results from PROs do not always track with the composite indices. When PROs are independent predictors of treatment response, PRO results should be reported separately rather than included in a composite score [33]. In one study, composite indices were better at detecting flare states though worse at describing states of low disease activity [34]. This supports the inclusion of PROs in clinical studies along with established composite indices.

The same investigators also studied disparate perspectives about the change in clinical status between clinicians and patients with rheumatoid arthritis. In this study, patients required a greater improvement in their condition to show improved satisfaction with the change compared to physicians. Likewise, patients required less deterioration to show dissatisfaction with their condition compared to clinicians. Clinicians valued a change in the disease activity score of equal magnitude – better or worse – to reflect improvement or deterioration. However, patients required more disease reduction to consider it an improvement and less worsening to trigger dissatisfaction. There was only about 60% concordance between patients and clinicians in judgment of disease activity. Perception of improvement/worsening has been shown in other studies to vary from the vantage point of clinician, patients, and caregivers [35]. This is an area needing additional study. Examples of discordance can vary depending on type of study, measures used, and how perspectives are elicited.

Composite measures upon which clinicians base their assessment of improvement or deterioration may not necessarily include key aspects of the disease that matter to the patient. For instance, in the DAS-28 (disease activity score) assessment of rheumatoid arthritis, 28 joints are assessed for tenderness or swelling, but this count does not include the feet or ankles. Therefore, if patients have foot involvement, which can be very painful, it is possible they may not be satisfied with the degree of improvement noted on the DAS-28 because it excludes an important element of their disease experience. Using the DAS-28 as an outcome in pharmacoepidemiology studies could therefore miss outcomes important to patients and result in biased measures of association. Further, composite disease activity measures may not always translate to decision making based on the experience of an individual patient [36]. As a matter of patient engagement, the viewpoint of the patient must be considered in collaborative treatment decisions, and it may not always be aligned with outcomes assessed in research studies [37].

Measuring Within-Person Change

Understanding how to estimate clinically meaningful change using PROs is important to be able to determine the effectiveness of a treatment or intervention. Determining clinically meaningful change is complex; there is no consensus on the best approach and the topic represents an area of active methodological research. Although there are standard statistical

approaches to understanding and analyzing the change in PRO scores over time, detection of a statistically significant difference may not reflect a meaningful clinical difference. As noted above, there may be differences in defining a meaningful clinical difference depending on the respondent (e.g., patient, caregiver, clinician). It may depend on the health condition being studied, and also whether a patient is experiencing improvement or deterioration at the time of measurement. Additional considerations that make classification of meaningful clinical change challenging include instability of intrarater judgment, temporal changes related to disease stage, progress, and severity (see Chapter 37). Perception, or recall, of progress compared to an earlier state of disease may also be subject to bias. The context of measurement or the consequence of declaring a change meaningful (e.g., resulting in a treatment change) could influence the study results and decisions about clinical approaches (see Chapter 39).

Statistical Methods

Patient-reported outcome scores may change in response to an intervention, but just because a PRO is responsive and a change in score is statistically detectable does not mean the change is important. The *minimal important difference* (MID) represents the smallest change in score that could be determined important [12]. The *minimal clinically important difference* (MCID) is determined based on clinical anchors [12], and is generally regarded as the smallest difference in score that would prompt a change in patient management [38]. In practice, the terms MID and MCID are sometimes used interchangeably, but there may be differences between statistically derived estimations and those anchored on clinical parameters. MID and MCID estimates vary according to the population being studied and context of measurement. The MID is a useful calculation to help determine responsiveness to change of a PRO measure, which is part of construct validity.

Anchor-based methods such as the MCID use external indicators considered to be clinically relevant to the PRO, such as clinical measures (lab tests, clinician ratings) or patient measures (global rating of change), and place subjects on a continuum based on the size of change in the anchor (large negative change, small negative change, no change, small positive change, large positive change). Estimating what is considered a small versus medium, or other, change on a given clinical measure may require a consensus process [39]. Ideally, multiple relevant anchors should be used across multiple samples to confirm responsiveness of the PRO measure. Patients who made small changes that were considered meaningful should be included in the exercise [12] to try and reduce the likelihood of picking statistical but not clinically important differences. Another anchor-based technique to estimate MID uses receiver operating characteristic (ROC) curves to evaluate group-level criteria for improvement or worsening of clinical status [40,41].

Distribution-based methods are more strictly based on statistics, without requirement for a clinical anchor. The distribution-based approach uses scores from a sample to express the effect in terms of standard deviation units or standard error of measurement [39]. Distribution methods can be used if anchor methods are not available, or to assess if anchor-based measures are reasonable estimates [12].

Bookmarking and Scale Judgment of IRT-Based Measures

Alternative approaches to measuring meaningful within-patient change have been developed based on techniques from the field of educational testing applicable to PRO measures developed using IRT methods. The general approach is for the development of clinical vignettes representing a continuum of IRT-based scores, presenting these vignettes to a representative panel of stakeholders (e.g., patients, caregivers, clinicians) and having the

panel identify thresholds between scores (delineated by the vignettes), where they place a "bookmark" separating different levels of severity. Imagine, for example, the case of physical function; vignettes would be presented of patients having no physical limitation through severe disability. The panel would order the cards according to physical function described by the vignette, then set thresholds between categories of patients with "no problem," "mild problem," "moderate problem," and "severe problem." Such exercises help to identify clinically meaningful cut-points between scores [42]. Similar qualitative work with panelists can be used to identify minimal clinically meaningful differences by presenting PRO items to stakeholders and asking them to note how much response to an item (or items on a scale) would need to change for a change in status to be considered clinically meaningful [35].

Another approach, the "scale judgment" method, entails raters comparing prefilled IRT-based PRO questionnaires (for example, considering before and after an intervention) and indicating whether or not the person who completed the questionnaires had experienced an important difference [43]. These types of approaches require qualitative work with patient and other stakeholders across different patient populations (age, demographics, health conditions) and considering direction of improvement and worsening of PROs. Research is needed into the impact of different stakeholder roles and perspectives on agreement on severity thresholds and MCIDs, and how to understand and reconcile discordance particularly in the arena of clinical decision making.

Change in perspective

Another factor which may complicate the assessment of within-person health state change is the situation when a person's perception and valuation of the domain being measured change over time. This phenomenon, termed *response shift*, is defined as:

> a change in the meaning of one's self-evaluation of a target construct as a result of: a) a change in the respondent's internal standards of measurement (scale recalibration); b) a change in the respondent's values (i.e., the importance of component domains constituting the target construct), or c) a redefinition of the target construct (i.e, reconceptualization) [44].

For example, this could occur if a psychological intervention resulted in a change in thinking about a person's condition and the impact on their life or if patients adapt to disease progression, get used to medication toxicities, etc. Response shifts could potentially alter estimates of treatment effect over time. There are both statistical and qualitative methods to assess for response shift, and guidelines for investigators to evaluate objective change and changes in internal valuation [44]. The phenomenon of response shift in the quality of life appraisal process of the patient respondent may also be at root when their global rating of change differs from what might be expected based on PROs or other assessments [45].

Currently Available Solutions

Practical Guides for Selection of PROs and Implementation into Practice

With the proliferation in interest and adoption of PROs, there are publications offering guidance on training clinicians in use of PROs, as well as a number of studies on presentation of PRO scores to patients and clinicians. Recommendations for training clinicians on PRO use include eliciting local barriers and concerns to address during the training (such as how to deal with patient symptoms/concerns outside the specialty, concern about visit time constraints, how to interpret results), inclusion of the stakeholders in PRO selection and format for presentation, keeping the

training relatively brief, making training problem based and experiential with video examples and case studies, and including relevant treatment decision aids and decision support tools in the training [46].

Experience on considerations for graphical display and communication of PRO scores has also been published. Display format preferences tend to vary by audience characteristics such as age, education, and role (clinician vs patient) [47]. Some studies have identified preferences for line graphs [48], others for bar graphs [49]. Patients tend to prefer simpler formats, and clinicians prefer more data [47,48]. Directionality of data has been shown to matter, with better health being portrayed as higher on the chart, and including lines indicating threshold values for normal versus abnormal found to be helpful [50]. There are published examples, and depending on EHR vendor there may be preprogrammed displays. As the field matures, additional resources and guidance are anticipated to become available.

Overcoming perceived barriers to PRO adoption in clinical care may take a cultural shift, including acceptance of co-production as a means to transform healthcare delivery and improve chronic illness care. For some, it may help to view PROs as a tool in the data arsenal for enhanced signal detection of the health of an individual gone awry, or as a communication tool for more efficient and accurate exchange of information. As evidence mounts on the side of increased quality of care and improved outcomes resulting from interventions supporting co-production, the tide may shift to embrace PROs use in clinical care.

There are a host of useful publications providing guidance for integrating PROs into clinical practice, including considerations for selection of PROs use in performance measurement, and guidance on integration of PROs into electronic health records. Example references are publicly available and the reader is also referred to the following resources:

- *User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice.* International Society for Quality of Life Research (prepared by Aaronson L, Elliott T, Greenhalgh J, Halyard M, Hess R, Miller D, Reeve B, Santana M, Snyder C). www.isoqol.org/UserFiles/2015UsersGuide-Version2.pdf (accessed May 7, 2019).
- *Patient Reported Outcomes (PROs) in Performance Measurement.* National Quality Forum. www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx (accessed May 7, 2019).
- *User's Guide to Integrating Patient Reported Outcomes in Electronic Health Records.* Prepared for the Patient Centered Outcomes Research Institute (PCORI) by Johns Hopkins University. Snyder C, Wu AW, eds. www.pcori.org/document/users-guide-integrating-patient-reported-outcomes-electronic-health-records (accessed May 7, 2019).
- A useful reference to find information about publicly available PRO measures developed and evaluated with National Institutes of Health (NIH) funding is the Health Measures website. This is a repository of official information about PROMIS® (Patient-Reported Outcomes Measurement Information System), which has a suite of person-centered measures across multiple domains of physical, mental, and social health for adult and pediatric populations. This repository serves as a distribution center for other NIH-funded measures such as Neuro-QoL, NIH Toolbox®, and ASCQ-Me®, in addition to PROMIS. www.healthmeasures.net/explore-measurement-systems/promis

The Future

Patient engagement in research, advances in PRO measurement, and recognition of the importance of garnering direct patient input will result in increased inclusion of the patient voice

in the calculus of medication efficacy in clinical studies [4–6]. PROs are increasingly used at the point of clinical care, and there are now prototypes for capturing data between clinical or study visits to enhance co-production of care [28]. There is growing evidence that electronic PROs, or at least symptom monitoring, may increase not only quality of care of processes, but also outcomes [23], and use of PROs may become a reflection of, and an instrument for, improving quality of care. PROs may become increasingly used in comparative effectiveness research, with patient engagement in study design leading to more meaningful incorporation of PRO measures. PRO capture is being incorporated in learning networks and registries [51]. As more is understood about PRO development, effective use, interpretation and potential applications, the use and new use cases for PROs can be expected to continue to grow.

There is increasing interest in using PRO data from clinical settings and captured in electronic health records as structured outcomes assessment for inclusion in comparative effectiveness research. This could be a powerful data source when combined with other sources of electronic data [52,53] (see Part IIIb). PROs in EHRs may be of particular use when there are no standardized outcomes assessments provided by clinicians in the clinical note. The complexities of analysis and interpretation of longitudinal PRO data require continued study to best leverage such data to make valid inferences. Cross-cutting PRO measures that could be used across conditions and contexts may confer advantages such as anticipating from clinical trials results the expected outcomes in clinical practice. Such PROs when collected in a clinical setting – or with technology to support in-between visit data collection – could be used such that clinical data registries could be combined with administrative claims data to support comparative effectiveness research.

An area of future development in medicine applicable to study and use of PROs is the use of wearable devices with health monitors. Previously used as sports fitness trackers, there is now the ability to track heart rhythms with electrocardiograms transmitted to medical facilities, and blood glucose monitoring may eventually be an option. Medical device companies and technology are making substantial investments in this arena. Coupling physiologic data with PRO data will be informative in understanding clinically relevant change in scores. For example, a patient may be able to better calibrate a worsening feeling of fatigue (quantified in PRO score) to corresponding physiologic worsening, keeping individual patients more in tune with how symptoms and function may reflect underlying physiology.

There are open research questions related to analysis and interpretation of measures used longitudinally. There are interesting and exciting case examples, and it remains to be seen which model will be scalable and generalizable to more settings. Ideally, the culture of co-production, patient engagement, and self-management will continue to take root, perhaps strengthened by demonstration of outcomes improvement, and support the shift towards PRO measurement for meaningful application, evidence generation, and shared decision making.

Acknowledgment

The author thanks Wendy Engstrom Gerhart, Director, Quality Outcomes and Evidence, Cincinnati Children's Hospital, as a colleague in implementation of patient-reported outcomes measures in clinical care and for her critical review and edits to the chapter.

References

1 Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press, 2001.

2 Batalden M, Batalden P, Margolis P, *et al*. Coproduction of healthcare service. *BMJ Qual Saf* 2016; **25**(7): 509–17.

3 Wagner EH, Austin BT, Davis C. Improving chronic illness care: translating evidence into action. *Health Aff* 2001; **20**: 64–78.

4 Forsythe L, Heckert A, Margolis M, *et al*. Methods and impact of engagement in research, from theory to practice and back again: early findings from the Patient-Centered Outcomes Research Institute. *Qual Life Res* 2018; **27**(1): 17–31.

5 US Department of Health and Human Services, FDA Center for Drug Evaluaton and Research, Research Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for Industry. Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. *Health Qual Life Outcomes* 2006; **4**: 79.

6 Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007; **16**(Suppl 1): 133–41.

7 Reeve BB, Hays RD, Bjorner JB, *et al*. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007; **45**(5 Suppl 1): S22–31.

8 Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006; **12**(5): 559–68.

9 Food and Drug Administration, BEST (Biomarkers, EndpointS, and other Tools). Glossary. www.ncbi.nlm.nih.gov/books/NBK338448/ (accessed May 5, 2019).

10 Duke-Margolis Center for Health Policy. Clinical Outcome Assessments: Establishing and Interpreting Meaningful Within-Patient Change. https://healthpolicy.duke.edu/events/clinical-outcome-assessments-establishing-and-interpreting-meaningful-within-patient-change (accessed May 5, 2019).

11 FDA. PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2013 through 2017. www.fda.gov/media/81306/download (accessed May 5, 2019).

12 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; **61**(2): 102–9.

13 Jayadevappa R, Cook R, Chhatre S. Minimal important differences to infer changes in health-related quality of life – a systematic review. *J Clin Epidemiol* 2017; **89**: 188–98.

14 Lavallee DC, Chenok KE, Love RM, *et al*. Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff* 2016; **35**(4): 575–82.

15 Frost MH, Bonomi AE, Cappelleri JC, *et al*. Applying quality-of-life data formally and systematically into clinical practice. *Mayo Clin Proc* 2007; **82**(10): 1214–28.

16 Patient Centered Outcomes Research Institute. User's Guide to Integrating Patient Reported Outcomes in Electronic Health Records. www.pcori.org/document/users-guide-integrating-patient-reported-outcomes-electronic-health-records (accessed May 5, 2019).

17 International Society for Quality of Life Research. User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice. https://pdfs.semanticscholar.

org/9f86/43e764e9ef42dd9715312e0f91d631a530f7.pdf?_ga=2.266317967.1436055729.1557073309-1091180841.1557073309 (accessed May 5, 2019).
18 Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract* 1999; **5**(4): 401–16.
19 Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. *Med Care* 2000; **38**(2): 175–86.
20 Nader CM, Tsevat J, Justice AC, *et al*. Development of an electronic medical record-based clinical decision support tool to improve HIV symptom management. *AIDS Patient Care STDS* 2009; **23**(7): 521–9.
21 Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient–physician communication: a randomized controlled trial. *JAMA* 2002; **288**(23): 3027–34.
22 Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: a structured review of intervention strategies. *Patient Educ Couns* 2006; **63**(1–2): 12–23.
23 Basch E. Patient-reported outcomes – harnessing patients' voices to improve clinical care. *N Engl J Med* 2017; **376**(2): 105–8.
24 Basch E, Deal AM, Kris MG, *et al*. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol* 2016; **34**(6): 557–65.
25 National Quality Forum. Patient Reported Outcomes (PROs) in Performance Measurement. file:///C:/Users/Owner/Downloads/PROFinalReport.pdf (accessed May 5, 2019).
26 Rotenstein LS, Huckman RS, Wagle NW. Making patients and doctors happier – the potential of patient-reported outcomes. *N Engl J Med* 2017; **377**(14): 1309–12.
27 Glasgow RE, Funnell MM, Bonomi AE, Davis C, Beckham V, Wagner EH. Self-management aspects of the improving chronic illness care breakthrough series: implementation with diabetes and heart failure teams. *Ann Behav Med* 2002; **24**(2): 80–7.
28 Kaplan HC, Thakkar SN, Burns L, *et al*. Protocol of a pilot study of technology-enabled coproduction in pediatric chronic illness care. *JMIR Res Protoc* 2017; **6**(4): e71.
29 Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care* 2008; **14**(8): 530–9.
30 Fung CH, Hays RD. Prospects and challenges in using patient-reported outcomes in clinical practice. *Qual Life Res* 2008; **17**(10): 1297–302.
31 Snyder CF, Herman JM, White SM, *et al*. When using patient-reported outcomes in clinical practice, the measure matters: a randomized controlled trial. *J Oncol Pract* 2014; **10**(5): e299–306.
32 Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *JAMA* 2012; **307**(15): 1583–4.
33 Orbai AM, Bingham CO 3rd. Patient reported outcomes in rheumatoid arthritis clinical trials. *Curr Rheumatol Rep* 2015; **17**(4): 28.
34 Leeb BF, Sautner J, Mai HT, Haindl PM, Deutsch C, Rintelen B. A comparison of patient questionnaires and composite indexes in routine care of rheumatoid arthritis patients. *Joint Bone Spine* 2009; **76**(6): 658–64.
35 Morgan EM, Mara CA, Huang B, *et al*. Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS(R)) measures in juvenile idiopathic arthritis using standard setting with patients,

parents, and providers. *Qual Life Res* 2017; **26**(3): 565–86.

36 Leeb BF, Sautner J, Leeb BA, Fassl C, Rintelen B. Lack of agreement between patients' and physicians' perspectives of rheumatoid arthritis disease activity changes. *Scand J Rheumatol* 2006; **35**(6): 441–6.

37 Hendrikx J, Kievit W, Fransen J, van Riel PL. The influence of patient perceptions of disease on medication intensification in daily practice. *Rheumatology* 2016; **55**(11): 1938–45.

38 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**(4): 407–15.

39 Guyatt GH, Osoba D, Wu AW, *et al*. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; **77**(4): 371–83.

40 Ward MM, Marx AS, Barry NN. Identification of clinically important changes in health status using receiver operating characteristic curves. *J Clin Epidemiol* 2000; **53**(3): 279–84.

41 Turner D, Schünemann HJ, Griffith LE, *et al*. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol* 2009; **62**(4): 374–9.

42 Cook KF, Victorson DE, Cella D, Schalet BD, Miller D. Creating meaningful cut-scores for Neuro-QOL measures of fatigue, physical functioning, and sleep disturbance using standard setting with patients and providers. *Qual Life Res* 2015; **24**(3): 575–89.

43 Thissen D, Liu Y, Magnus B, *et al*. Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method. *Qual Life Res* 2016; **25**(1): 13–23.

44 Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med* 1999; **48**(11): 1531–48.

45 Schwartz CE, Powell VE, Rapkin BD. When global rating of change contradicts observed change: examining appraisal processes underlying paradoxical responses over time. *Qual Life Res* 2017; **26**(4): 847–57.

46 Santana MJ, Haverman L, Absolom K, *et al*. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. *Qual Life Res* 2015; **24**(7): 1707–18.

47 Bantug ET, Coles T, Smith KC, *et al*. Graphical displays of patient-reported outcomes (PRO) for use in clinical practice: what makes a PRO picture worth a thousand words? *Patient Educ Couns* 2016; **99**(4): 483–90.

48 Brundage MD, Smith KC, Little EA, *et al*. Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. *Qual Life Res* 2015; **24**(10): 2457–72.

49 Izard J, Hartzler A, Avery DI, Shih C, Dalkin BL, Gore JL. User-centered design of quality of life reports for clinical care of patients with prostate cancer. *Surgery* 2014; **155**(5): 789–96.

50 Snyder CF, Smith KC, Bantug ET, *et al*. What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. *Cancer* 2017; **123**(10): 1848–59.

51 Devine EB, Alfonso-Cristancho R, Devlin A, *et al*. A model for incorporating patient and stakeholder voices in a learning health care network: Washington State's Comparative Effectiveness Research Translation Network. *J Clin Epidemiol* 2013; **66**(8 Suppl): S122–9.

52 Wu AW, Snyder C, Clancy CM, Steinwachs DM. Adding the patient perspective to comparative effectiveness research. *Health Aff* 2010; **29**(10): 1863–71.

53 Wu AW, Kharrazi H, Boulware LE, Snyder C. Measure once, cut twice – adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol* 2013; **66**(8 Suppl): S12–20.

43

Advanced Approaches to Controlling Confounding in Pharmacoepidemiologic Studies

Sebastian Schneeweiss[1] *and Samy Suissa*[2]

[1] *Harvard Medical School; Harvard TH Chan School of Public Health; Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA*
[2] *Department of Epidemiology, Biostatistics and Medicine, McGill University; Centre for Clinical Epidemiology, Lady Davis Research Institute – Jewish General Hospital, Montreal, Quebec, Canada*

The past two decades have witnessed an explosion of methodological advances in the design and analysis of epidemiologic studies. Some of these contributions have been fundamental to the field of epidemiology in general while others have arisen specifically from questions posed by pharmacoepidemiologic applications. Several of these advances have already played an important role in the conduct of research on drug effects, and will take an even greater place in future applications. In this chapter, we introduce some of these approaches with a focus on confounding control, one of the major methodologic challenges in drug safety and effectiveness research with noninterventional studies.

We start out by describing a robust study design that will exemplify several aspects of confounding control and other biases and point out critical decision points in the choice of study designs. Second, we describe efficient sampling strategies within cohort studies (case–control, case–cohort, and two-stage sampling) and self-controlled designs (case–crossover and case–time–control designs) and how they will help reduce confounding bias. Third, we introduce several analytic methods that have gained wider use in pharmacoepidemiologic studies and others that only recently have made inroads into pharmacoepidemiology.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Pharmacoepidemiologic analyses are in principle no different from analyses in any other subject area within epidemiology. They are concerned with valid estimation of associations between an exposure and outcome, and methods to minimize systematic and random error that may cloud causal conclusions. Some issues specific to pharmacoepidemiology stem from the constraints of the frequently used secondary data sources, in particular large electronic longitudinal healthcare databases from insurance health plans, electronic medical records systems, or registries (see Chapters 11–14). Another difference is the often unusually direct interdependency of treatment choice with health status, severity of disease, and prognosis that may lead to strong, sometimes intractable confounding by indication (see Chapter 3) [1].

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Pharmacoepidemiologists try to reduce biases by appropriate choices of study design and analytic strategies. Challenges arise if not all confounder information is captured in the available data. This chapter provides an overview of selected options that fit typical pharmacoepidemiologic data sources and study questions.

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

The ready and relatively cheap availability of large longitudinal patient-level healthcare databases make the new-user cohort design a natural design choice as a starting point that mimics the classic parallel group controlled trial, except of course for the randomized treatment assignment (Figure 43.1) [2]. Efficient sampling designs within such cohorts, including case–control, case–cohort, and two-stage sampling designs, are important extensions to assess additional covariate or outcome information in a subset of patients. Such sampling usually provides no advantage if secondary data are the only source for exposure, covariate, or outcome assessment because there is no additional cost or time to analyze the entire dataset rather than a subsample [3].

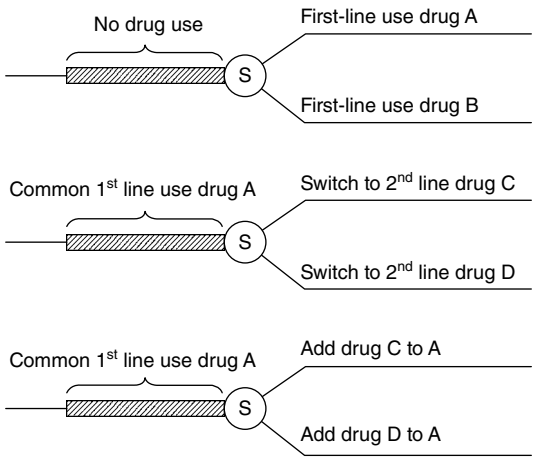


Figure 43.1 Principle of the new user design and its variations when studying second-line therapies. *Source:* Reproduced from Schneeweiss [3] with permission from John Wiley & Sons.

Bias can be reduced by appropriate design choices. Considerations about the sources for exposure variation will lead to fundamental decisions on the appropriate study design. In a causal experiment, one would theoretically expose a patient to an agent and observe the agent's effect on his or her health, then rewind time, leave the patient unexposed, and keep all other factors constant to establish a counterfactual experience. Since this experiment is impossible, the next logical expansion of the experiment is to randomly introduce or observe exposure variation within the same patient but over time (Figure 43.2). If we observe sporadic drug use resulting in fluctuations of exposure status within a patient over time, if that drug has a short washout period, and if the adverse event of interest has a rapid onset, then we may consider a case-crossover design or related approaches (see later). For most pharmacoepidemiologic studies, we will utilize variation in exposure between individual patients, and we will therefore apply a cohort study design. Any exposure variation among higher-level entities (provider, region, etc.) can be exploited using instrumental variable analyses (described later in the chapter) if unrelated to patient characteristics either directly or indirectly [4].

In a cohort design, there are several advantages to identifying patients who start a new drug and begin follow-up after initiation, similar to a parallel group randomized controlled trial that establishes inception cohorts. As patients in both the study group and the comparison group have been newly started on medications, they have been evaluated by physicians who concluded that these patients might benefit from the newly prescribed drug. This makes treatment groups more similar in characteristics that might not be observable in the study database if medication use is not new.

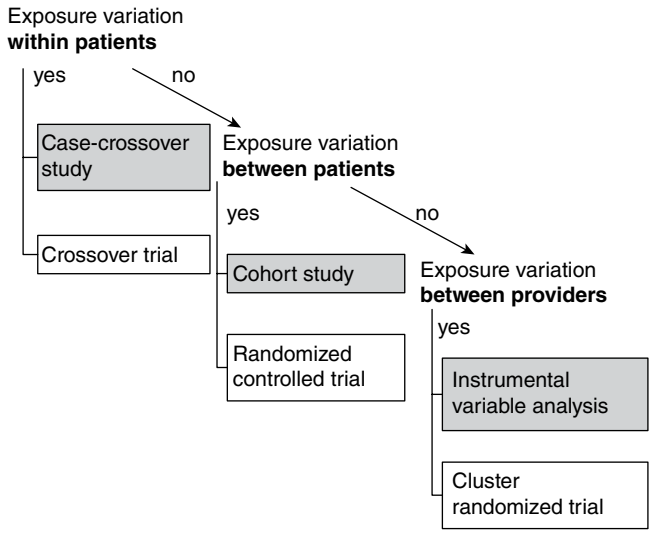


Figure 43.2 Study design choice by source of exposure variation. Shaded boxes indicate noninterventional study designs while clear boxes are the randomized design versions. *Source:* Reproduced from Schneeweiss [3] with permission from John Wiley & Sons.

The clear temporal sequence of confounder measurement before treatment initiation in an incident user design also avoids mistakenly adjusting for consequences of treatment (intermediates) rather than predictors for treatment, a possible reason for "overadjustment" [5]. Comparing two active treatment groups further reduces the chances of immortal time bias, a mistake that most frequently emerges if future information is used to define earlier exposure status in healthcare databases, particularly when defining a "nonuser" comparison group [6]. A common example of immortal time bias is to define nonusers as patients who have not used the study medication during the first six months of follow-up. By definition, these nonuser patients cannot die during the first six months of follow-up, and therefore their inclusion can bias the findings. Because of the well-defined starting point of inception cohorts, it is possible to assess whether and in what form hazards vary over time by stratifying on duration of treatment. Studying new users is also useful when studying newly marketed medications; the incident user design avoids comparing populations predominantly composed of first-time users of a newly marketed drug with a population predominantly composed of prevalent users of the old drug. Such a comparison would be prone to bias because patients who stay on treatment for longer may be less susceptible to the event of interest [7].

A common criticism of the incident user design is that excluding prevalent users will restrict and thus reduce the study size, in some cases substantially [8]. While this is true, researchers should be aware that by including ongoing (prevalent) users, they might gain precision at the cost of validity [9]. Screening and identifying incident users in secondary databases is not costly except for a bit more computing time. In some situations, particularly studies of second-line treatments in chronic conditions, we can only study patients who switch from one drug to another, as very few patients will be treatment naive. Such switching is often not random, but rather is determined by progressing disease and treatment failure or by side effects that may be related to the study outcome. A fairer treatment comparison may be achieved by comparing new switchers to the study drug with new switchers to a comparison drug (see Figure 43.1). Consequently, prevalent new-user cohort designs are being developed to minimize bias when one needs to include as many new users of the study drug as possible.

Even with appropriate designs, however, all observational pharmacoepidemiologic studies still must consider carefully how to approach the problems of potential confounding, in order to prevent bias. Approaches to addressing these methodologic challenges, and their limitations, will be the primary focus of this chapter.

Currently Available Solutions

The solutions available to minimize confounding in pharmacoepidemiologic database studies can be broadly categorized into (1) approaches that collect more information on potential confounders and apply efficient sampling designs to reduce the time and resources it takes to complete the study, and (2) analytic approaches that try to make better use of the existing data with the goal of improved control of confounding.

Efficient Sampling Designs Within a Cohort Study

In any cohort study, the cost, time, and resources necessary to collect data on all cohort members can be prohibitive. Even with cohorts formed from computerized databases, there may be a need to supplement and validate data with information from hospital records, medical records, and physician or patient interview questionnaires, with the goal of improved confounding control. When the cohort size is considerable, such additional data gathering can become a formidable task. Moreover, even if no additional data are needed, the data analysis of a cohort with multiple and time-dependent drug exposures can become technically unfeasible, particularly if the cohort size and number of outcome events are large. For example, a study of the long-term effect of antihypertensive drugs and the risk of cancer involved a cohort of over 1.1 million patients where 41 059 developed cancer during 14 years of follow-up, a size that necessitated sampling within the cohort [10]. Finally, there are situations with multiple confounding factors that may require accurate matching rather than simply modeling adjustment.

To counter these constraints, designs based on sampling subjects within a cohort have been proposed and applied successfully in pharmacoepidemiology. These designs are based on the selection of all cases with the outcome event from the cohort, but differ in the selection of a small subset of "noncases." Generally, they permit the precise estimation of relative risk measures with negligible losses in precision. Below, we discuss structural aspects of cohorts and present three sampling designs within a cohort the nested case–control, the multi-time case–control, and case–cohort designs.

Structures of Cohorts

Figure 43.3 illustrates graphically a cohort of 21 newly diagnosed diabetics over the period 1995 to 2010. This cohort is plotted in terms of calendar time, with subjects ranked according to their date of entry into the cohort, which can correspond to the date of disease diagnosis or treatment initiation. Such *calendar-time cohorts* depict the natural chronological nature of the cohort accrual. An alternative depiction of this same cohort could be based on duration of disease (i.e., follow-up time from diagnosis or first

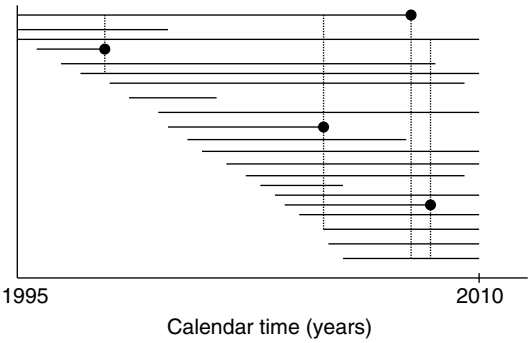


Figure 43.3 Illustration of a *calendar-time* cohort of 21 subjects followed from 1978 to 1990 with four cases (●) occurring and related risk-sets (---).

exposure to a drug), which may be more relevant to the drug effect under study. In this instance, the illustration given in Figure 43.4 for the same cohort, using follow-up time as the new time axis, is significantly different from the previous one. In these *follow-up-time cohorts*, the same subjects are ranked according to the length of follow-up time in the study with zero-time being the time of diagnosis or treatment start.

The question of which of the two forms one should use for the purposes of data analysis rests on one's judgment of the more relevant of the two time axes, essentially the one for which the risk varies most over time, called the primary time axis, with respect to risk and drug exposure. This decision is important, since it affects the demarcation of "risk-sets," which are fundamental to the analysis of data from cohorts and consequently the sampling designs within cohorts. A risk-set is formed by the members of the cohort who are at risk of the outcome event at a given point in time; namely they are free of the outcome event and are members of the cohort at that point in time, called the index date. Drug exposure measures are then anchored at this index date. It is clear that Figures 43.3 and 43.4 produce distinct risk-sets for the same cases in the same cohort, as illustrated by the different sets of subjects crossed by the vertical broken line for the same case under the two forms of the cohort. In Figure 43.3, for example, the first chronological case to occur has in its risk-set only the first six subjects to enter the cohort, while in Figure 43.4, all 21 cohort members belong to its risk-set at the time that the first case arises. While the second form based on disease duration is often used, because in pharmacoepidemiologic drug exposure can vary substantially over calendar time, the first form may be as relevant for the formation of risk-sets and data analysis as the second form. Regardless, an advantage of having data on the entire cohort is that the primary time axis can be changed according to the study question, using calendar time for one analysis, duration of disease or drug exposure for another, with respective adjustment in the analysis for the effect of the other time axis.

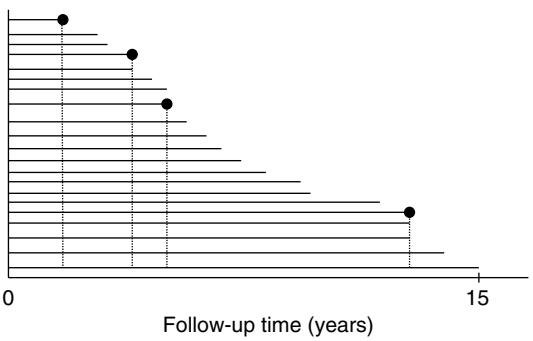


Figure 43.4 Illustration of *follow-up-time* cohort representation after rearranging the cohort in Figure 43.3, with the new risk-sets (---) for the four cases.

The Nested Case–Control Design

The notion of a nested case–control design within a cohort was first introduced by Mantel [11], who proposed an unmatched selection of controls and called it a synthetic retrospective study. It was developed further and formalized by Liddell *et al.* [12] in the context of a cohort study of asbestos exposure and the risks of lung cancer and mortality. The modern nested case–control design involves four steps:

1) defining the cohort time axis, as above
2) selecting all cases in the cohort, i.e., all subjects with an outcome event of interest
3) forming a risk-set for each case and
4) *randomly* selecting one or more controls from each risk-set.

Figure 43.5 illustrates the selection of a nested case–control sample from a cohort, with one control per case (1:1 matching). It is clear from the definition of risk-sets that a future case is eligible to be a control for a prior case, as illustrated in the figure for the fourth case (the circle occurring last in time), and that a subject may be selected as a control more than once. A bias

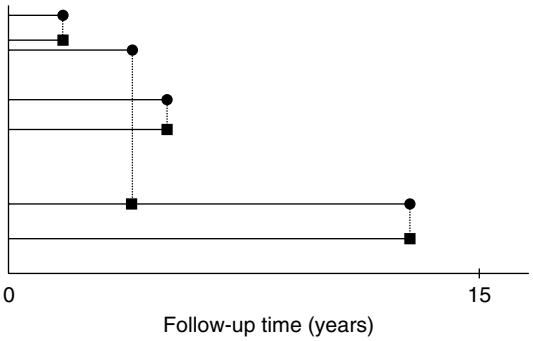


Figure 43.5 Nested case–control sample of one control (■) per case (●) from cohort in Figure 43.4.

is introduced in the estimation of the relative risk if controls are forced to be selected only from the noncases and subjects are not permitted to be used more than once in the nested case–control sample, since the control exposure prevalence will be slanted to that of longer term subjects who do not become cases during the study follow-up [13]. The magnitude of the bias depends on the frequency of the outcome event in the cohort; the more frequent the event, the larger the potential for bias.

This property leading to subjects possibly being selected more than once in the sample may be challenging when the exposure and covariate factors are time dependent, particularly when the data are obtained by questionnaire where the respondent would have to answer questions regarding multiple time points in their history. This issue arose in a study of the risks of severe adverse events in asthma associated with the use of inhaled beta-agonists [14]. A cohort of 12 301 asthmatics spanning the period 1978–87 was identified from the Saskatchewan Health computerized databases, of whom 129 were cases (death or near-death from asthma). A nested case–control approach was needed to permit the collection of additional data from hospital charts and questionnaires sent to all physicians who saw these patients. These additional data were time dependent, focusing on the two-year period prior to the index (risk-set) date. A standard nested case–control sample of six controls per case, as described above, would have likely produced some case and control subjects who contributed multiple times as controls in the sample. This would have added complexity to the questioned physicians who, for example, would have had to respond to questions about the same patient's asthma severity in different two-year periods, a potentially confusing data collection scheme. In part to circumvent this difficulty, the cohort was stratified according to various potential confounding factors, namely age, area of residence, social assistance, prior asthma hospitalization and calendar date of entry into the cohort. This fine stratification resulted in 129 mutually exclusive subcohorts, one leading to each case, and between two and eight controls per case (some risk-sets contained only two eligible controls). Since each subcohort contained a single risk-set (only one case) and the subcohorts were mutually exclusive, a selected subject was guaranteed to appear only once in the nested case–control sample.

The analysis of data from a nested case–control study must preserve the matched nature of the selection of cases and controls, particularly if the risk of the event changes with disease duration and drug exposure varies in calendar time. The method of analysis is identical to that of a conventional matched case–control study, not nested within a cohort. The conditional logistic regression method for this design is appropriate, as it uses the risk-set as the fundamental unit of analysis, in agreement with the proportional hazards model of the full cohort [15]. Simple formulae exist to estimate the relative risk for 1:1 matching [16].

The question of the required number of controls per case is important (see also Chapter 4). Although selecting one control per case will greatly simplify the data analysis, a large number of cases will be required to attain an acceptable level of power. Since the number of cases in the cohort is fixed and cannot be increased to

satisfy this requirement, the only remaining alternative is to increase the control-to-case ratio. Tables for determining the power for given numbers of controls are given in Breslow and Day [17], and Appendix A in this book. It can be readily seen from these sample size tables that the gain in power is significant for every additional control up to four controls per case, but becomes negligible beyond this ratio. For example, if we consider an exposure prevalence in the controls to be 30% and we target detecting a relative risk of 2 with 5% significance and 80% power, the required numbers of cases are 122, 90, 74, 65, and 62, respectively, for 1:1, 2:1, 4:1, 10:1, and 20:1 control-to-case ratios. These translate to total study sizes (cases and controls combined) of 244, 270, 370, 715, and 1302, with clear cost implications and related optimality decisions. Of course, the number of cases in a cohort is frequently fixed *a priori* by the study constraints, thus eliminating this option to increase the number of cases.

However, although this general rule of an optimal 4:1 control-to-case ratio is appropriate in the majority of instances, one should be prudent when exposure to the drug under study is infrequent, or when several factors or other drugs are being assessed simultaneously. In these situations, the ratio could easily be required to increase to 10 or more controls per case. This was the case in two recent nested case-control studies, within a cohort of over 40 000 patients with rheumatoid arthritis, where 100 controls per case were used to obtain sufficiently stable estimates of the rate ratios of serious hepatic events (n=25 cases) and interstitial lung disease (n=74 cases) associated with the use of disease-modifying antirheumatic drugs (DMARD) [18,19].

The appropriate method to perform external comparisons using data from a nested case–control design has been described [20]. It uses knowledge about the sampling structure to yield an unbiased estimate of the outcome event rate in the full cohort, thus permitting the estimation of the necessary standardized relative measure with respect to the selected external population.

Finally, the "nested case–control" label has led to some misunderstandings, including the usual presentation of data as a comparison between "cases" and "controls" rather than by exposure, as well as the convoluted way that forward-looking associations from exposure to outcome extracted from backward-looking data. Moreover, the nested case–control approach provides estimates of the odds ratio, not a rate difference. However, the fact that it is nested within a clearly defined cohort with known sampling fraction allows estimation of risks and rates [21]; the quasi-cohort approach utilizes this property to address these concerns [22]. A quasi-cohort approach was used to assess the risk of pneumonia associated with inhaled corticosteroids in patients with asthma [23].

The Multi-time Case–Control Design

The multi-time case–control design has been introduced recently as an alternative strategy to improve the precision of the odds ratio in a case–control study with transient time-varying exposures, in a setting where increasing the number of control subjects is too costly. This approach is based on increasing the number of observations per control subject, by measuring drug exposure at many different points in time. Indeed, several case–control studies will collect extensive data on time-dependent exposures but use only a portion of these data in estimating the rate ratio.

For example, the International Agranulocytosis and Aplastic Anemia Study (IAAAS) assessed the risk of agranulocytosis associated with the use of analgesics using a case–control study of 221 cases of agranulocytosis and 1425 controls [24]. While the study collected data on exposure for four weeks prior to the index date, only one week's worth of data was used in the analysis. The multi-time case–control approach allows the use of all available exposure data during the four weeks (i.e., four control person-moments)

rather than only one week (i.e., one control person-moment) to improve the precision of the odds ratio estimate, which must however be corrected for within-subject correlation.

This design increases the number of control observations per case, thus potentially also increasing the power of the study without adding additional subjects [25]. For example, in a nested case–control study within a cohort of 12 090 patients with chronic obstructive pulmonary disease (COPD), there were 245 incident cases of acute myocardial infarction (AMI) that occurred during follow-up, for whom one and 10 controls per case were identified [25]. The rate ratio of AMI associated with use of antibiotics in the month prior to the index date was 2.00 (95% confidence interval [CI] 1.16–3.44) with one control per case. The precision (as reflected in the confidence intervals) was improved by increasing to 10 controls per case with a rate ratio of 2.13 (95% CI 1.48–3.05). Alternatively, keeping only one control patient per case but increasing the number of control time windows per subject from one to 10 (taken as 10 control exposure measures, one for each of the 10 months prior to the index date) also improved the precision with a rate ratio of 1.99 (95% CI 1.36–2.90).

The Case–Cohort Design

The first recognized application of a sampling design we currently call case–cohort was made by Hutchison [26], in performing external comparisons of leukemia rates in patients treated by radiation for cervical cancer. It was ultimately developed and formalized by Prentice [27], who coined the name "case–cohort." Although recent, this design has already been used effectively in some drug risk studies [28–31]. The case–cohort design involves two steps:

1) selecting all cases in the cohort, i.e., all subjects with an adverse event; and
2) randomly selecting a sample of predetermined size of subjects from the cohort, irrespective of case or control status.

Figure 43.6 depicts the selection of a case–cohort sample of six subjects from the illustrative cohort. Note that it is possible that some cases selected in step 1 are also selected in the step 2 sample, as illustrated in the figure for the third case.

The case–cohort design resembles a reduced version of the cohort, with all cases from the full cohort included. It can also be perceived as an unmatched version of the nested case–control design, with all cases compared with a random sample of the cohort used as controls though not at a specific person-moment. Although these aspects suggest a possible resemblance of the data analysis approach with either the established cohort or case–control methods, the techniques are in fact distinct, each requiring specific software. The analysis of case–cohort sampled data is complex, as it must take into account the overlap of cohort members between successive risk-sets induced by this sampling strategy [32].

The first advantage of the case–cohort design is its capacity to use the same sample to study several different types of events. Indeed, the cases can be split into several subcategories and each can be analyzed with the same "control" subcohort [33]. In contrast, the nested case–control design requires different control groups for each type of event because the selection

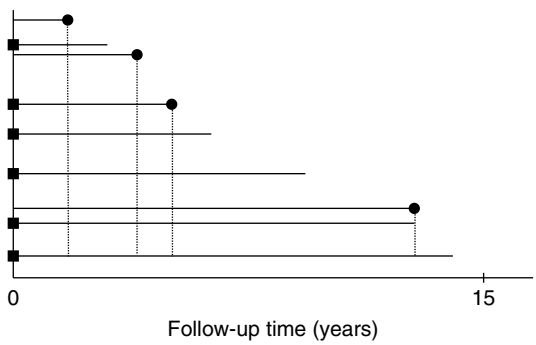


Figure 43.6 Case–cohort sample with six controls (■) from cohort in Figure 43.4.

depends on event times. For example, the beta-agonist risks nested case–control study had two distinct control groups, one of size 233 for the 44 asthma deaths, the other of size 422 for the 85 asthma near-deaths [14]. Another useful advantage is that the case–cohort design permits one to change the primary time axis of analysis from calendar to disease time and vice versa, depending on either the assumed model or the targeted outcome. This is not possible with the nested case–control study, where the primary time axis must be set *a priori* to permit the risk-set construction. This is less of a problem in pharmacoepidemiology, however, where the cohort can be divided into subcohorts of successive calendar time, as was discussed earlier. Yet another example is its simplicity in sampling, which has advantages in both comprehensibility and computer programming. Finally, external comparisons are simple to perform with the case–cohort approach [34].

The nested case–control design does have some advantages over the case-cohort design. The first is the simplicity of power calculation, or equivalently sample size determination. The nested case–control design is independent of the size of the cohort, while for the case–cohort design knowledge about overlap in risk-sets is essential, thus greatly complicating these calculations. Second, data on time-dependent exposure and covariates need only be collected up to the time of the risk-set for the nested case–control study, while the collection must be exhaustive for the case–cohort. Finally, despite the accessibility of software for data analysis of case–cohort data, these can quickly become surpassed and even infeasible with some of the huge sample sizes in some databases and multiple time-dependent exposures. In this situation, the nested case–control design, with its single risk-set per case, is not only advantageous but also the only solution. A study of benzodiazepine use and motor vehicle crashes, initially designed as a case–cohort study, had to be analyzed as a nested case–control study because of technical limitations of the case–cohort analysis software and hardware [35].

An obvious practical consideration is that the case–cohort sampling design can be used to study multiple endpoints in a single analysis (in contrast to case–control sampling) while the case–control study can easily consider many exposures. Depending on the clinical context, one might have strong preferences. As pointed out earlier, in database analyses the main use of cohort sampling designs is when additional information is needed that is time consuming or expensive to collect. If, for example, one engages in outcome validation via expensive chart review, a case–control analysis is often embedded in a cohort study [36]. On the other hand, if baseline biomarkers need to be obtained to determine patient subgroups or improve confounding control the case–cohort design is more suitable [37].

Prevalent New-User Designs

A common situation in pharmacoepidemiology involves the study of the effect of a new drug entering the market, with the best comparator being an older drug. Most often, patients prescribed the new drug will have been switched from the older comparator drug. An incident new-user cohort design based on incident new users of the study and comparator drugs, including only patients who were treatment naive to both drugs, would be optimal. However, it would exclude the possibly large number of subjects who switched from the older to the new drug, a clinically relevant subset. The prevalent new-user design provides an approach to include these switchers [38].

A prevalent new-user cohort is formed from the base cohort of all users of the comparator drug and of the drug under study, which inherently includes the subjects who switched from the comparator to the study drug those who initiated the study drug *de novo*. These latter subjects can directly be matched to contemporaneous initiators on covariates or propensity scores (see

below). For the subjects who switched from the comparator to the study drug, comparators can be selected from the base cohort by matching conditional on exposure sets. Time-based exposure sets can be defined, within the base cohort, by the time from the first prescription of the comparator drug up to the point of switching, while prescription-based exposure set are defined by the number of prescriptions of the comparator drug received up to the point of switching. Because of the granularity of the time scale, time-based exposure sets must be defined with a time interval (such as ±1 month) where all patients with a comparator prescription in the time interval belong to the exposure set and the set is defined by each patient's prescription date. Thus, with either type of exposure set, each switcher to the study drug will belong to an exposure set that includes subjects of similar duration or prescription history with a dispensing of the comparator drug. The importance of the exposure sets is that a visit occurred where the physician decided to either continue the comparator treatment or switch to the new study drug. The exposure set provides equivalent time points in the disease course at which confounding patient characteristics can be measured and controlled for.

To identify, within the exposure sets, the comparator drug users most similar to the patients who switched to the study drug, time-conditional propensity scores (TCPS) can be used [38]. The time-dependent Cox proportional hazards model can be used to compute the "propensity" of switching to the study drug, versus continuing on the comparator drug, as a function of the time-varying patient characteristics measured at the point of the exposure set, thus conserving the matching induced by the exposure set and avoiding adjusting for causal intermediates. The model is used to compute the time-conditional propensity scores within each exposure set, thus identifying their matched comparator as the one with the closest value to that of the switcher. For the purposes of the positivity assumption, the time-conditional propensity score of the switcher should lie within the range of the time-conditional propensity scores of the members of the corresponding exposure set. To emulate the randomized trial, the selection process can be initiated with the first chronological index study drug subject and repeated sequentially. Additionally, once a patient has been selected into the comparator group, they are not considered any longer in subsequent exposure sets as potential comparators. Thus, each subject who initiated the study drug will have a comparator user, matched on the time-conditional propensity score. Cohort entry is taken as the date of the first prescription of the study drug and the corresponding date for the matched comparator. If the exposure sets are too large to compute the time-conditional propensity scores by the time-varying Cox model, an alternative is to select random samples of prescriptions from each exposure set using conditional logistic regression, matching on the exposure set, with the relative odds estimating the relative hazards. The computed propensity odds score for the index switcher is used to identify the corresponding matched patient as the subject with the closest value from all members of the exposure set, not only the sampled ones.

This approach is useful for studies having a "nonuse" comparator, by using a physician visit or prescription for any drug other than the study drug as the comparator. Several questions remain regarding this design [38]. In particular, potential bias from using the prevalent users as comparators should be investigated by stratification on the incident/prevalent new-user status.

Within-Subject Designs

When dealing with the study of transient drug effects on the risk of acute adverse events, Maclure asserts that the best representatives of the source population that produced the cases would be the case subjects themselves; this is

the premise of the case–crossover design [39]. This is a design where comparisons between exposures are made within subjects, thus significantly attenuating the problem of confounding. An extension to the case–crossover design, the case–time–control design, has been proposed and is also presented here.

Case-Crossover Design

The case–crossover study is simply a crossover study *in the cases only*. The subjects alternate at varying frequencies between exposure and nonexposure to the drug of interest, until the adverse event occurs, which happens for all subjects in the study sample, since all are cases by definition. With respect to the timing of the adverse event, each case is investigated to determine whether exposure was present within the predetermined effect period, namely within the previous four hours in our example. This occurrence is then classified as having arisen either under drug exposure or nonexposure on the basis of the effect period. Thus, for each case, we have either an exposed or unexposed status, which represents for data analysis the first column of a 2 × 2 table, one for each case. Since each case will be matched to itself for comparison, the analysis is matched and thus we must create separate 2 × 2 tables for each case.

With respect to control information, the data on the average drug use pattern are necessary to determine the typical probability of exposure during the time window of effect. This is done by obtaining data for a sufficiently stable period of time. In our example, we may find out the average number of times a day each case has been using beta-agonists (two inhalations of 100 μg each) in the past year. Note that there are six four-hour periods (the duration of the effect period) in a day. Such data will determine the proportion of time that each asthmatic is usually spending time in the effect period and thus potentially "at risk" of ventricular tachycardia. This proportion is then used to obtain the number of cases expected on the basis of time spent in these "at-risk" periods, for comparison with the number of cases observed during such periods. This is done by forming a 2 × 2 table for each case, with the corresponding control data as defined above, and combining the tables using the Mantel–Haenszel technique as described in detail by Maclure [39].

To carry out a case–crossover study, three critical points must be considered. First, the study must necessarily be dealing with an acute adverse event that is alleged to be the result of a transient drug effect. Thus, drugs with chronic or regular patterns of use which vary only minimally between and within individuals are not easily amenable to this design. Nor are latent adverse events, which only occur long after exposure. Second, since a transient effect is under study, the effect period (or time window of effect) must be precisely determined. For example, in a study of the possible acute cardiotoxicity of inhaled beta-agonists in asthmatics, this effect period can be determined to be four hours after having taken the usual dose of two inhalations of 100 μg of the product. An incorrect specification of this time window can have important repercussions on the risk estimate, as we will show in the example below. Third, one must be able to obtain reliable data on the usual pattern of drug exposure for each case, over a sufficiently long period of time (as discussed further below). For our example, we could seek the frequency of use of beta-agonists during the year preceding the adverse event.

We generated data for a hypothetical case–crossover study of 10 asthmatic patients who experienced ventricular tachycardia. These were all queried (also hypothetically) regarding their use of two puffs of inhaled beta-agonist in the last four hours and on average over the past year. The data are displayed in Table 43.1. The fact of drug use within the effect period for the event classification is straightforward. The usual frequency of drug use per year is converted to a ratio of the number of "at-risk" periods to the number of "no-risk" periods, the total number of

four-hour periods being 2190 in one year. Thus, for example, the content of the 2 × 2 table for the first case, who is not found to have been exposed in the prior four-hour period, is (0,1,365,1825), while for the second case, who is exposed, it is (1,0,6,2184). Using the Mantel–Haenszel technique to combine the 10 2 × 2 tables, the estimate of relative risk is 3.0 (95% CI 1.2–7.6).

This method is sensitive to the specification of the time window of effect. For example, if this effect period is in fact only two hours, then the data of Table 43.1 would be affected in two ways: some cases may not be considered exposed any more, and the exposure probabilities will change. By considering as unexposed cases 2 and 4, for instance, who may have been exposed three hours before ventricular tachycardia, and recalculating the appropriate exposure probabilities, the relative risk becomes 2.0 (95% CI 0.3–12.0). On the other hand, if this effect period is in fact six hours long, then the data of Table 43.1 would be affected in two ways: some new cases may now be considered exposed, and the exposure probabilities will change. By considering as exposed cases 3 and 5, for instance, who may have been exposed five hours before ventricular tachycardia, and recalculating the appropriate exposure probabilities, the relative risk becomes 5.0 (95% CI 2.0–12.2). The difference in the magnitude of the risk and the corresponding statistical significance between the various scenarios is indicative of the importance of the need for an accurate specification of the length of the effect period.

This method is valuable when studying an acute adverse event that is alleged to be the result of a transient drug effect. Consequently, it excludes the study of drugs with regular patterns of use that vary minimally within individuals or adverse events which can only result from long extended exposure. Moreover, the case–crossover design requires precise knowledge about the effect period (or time window of effect), although the latter can be varied to investigate the optimum window to use. The design is also very useful when the selection of controls in the usual sense is uncertain. A significant advantage of this design is that it elimi-

Table 43.1 Hypothetical data for 10 subjects with ventricular tachycardia included in a case–crossover study of the risk of ventricular tachycardia in asthma associated with the four-hour period after beta-agonist exposure.

Case #	Beta-agonist use in last 4 hours[a] (E_i)	Usual beta-agonist use in last year	Periods of exposure (N_{1i})	Periods of no exposure (N_{0i})
1	0	1/day	365	1825
2	1	6/year	6	2184
3	0	2/day	730	1460
4	1	1/month	12	2178
5	0	4/week	208	1982
6	0	1/week	52	2138
7	0	1/month	12	2178
8	1	2/month	24	2166
9	0	2/day	730	1460
10	0	2/week	104	2086

[a] Inhalations of 200 µg: 1 = yes, 0 = no.

Note: Rate ratio estimator is $(\Sigma\, E_i\, N_{0i}) / (\Sigma\, (1- E_i)\, N_{1i})$.

nates the problem of confounding by factors that do not change over time. It cannot, however, easily address the problem of confounding by factors that do change over time. In this instance, time-dependent data will be required for such confounders, a possibly difficult task.

The case–crossover design is automatically free of control selection bias, which occurs when controls are not representative of the base population from which the cases arose. However, the case was inevitably different during the time period when they took the drug, from the time period when they did not take the drug. Thus, in this design, confounding by indication (see Chapter 33) can be severe. Although such control selection bias (in the usual control sense) is eliminated, case selection bias could be present if case selection is related to the exposure under study. Information bias resulting from the differential quality of recent and past drug exposure data can be a concern but less so if one uses drug exposure data from computerized databases. However, this design requires very precise knowledge of when a drug was actually taken, often a very difficult task in computerized databases, especially with drugs that are taken intermittently, exactly when this design could be useful.

Finally, the case–crossover design is intended to be used with transient exposures; otherwise estimates will be biased towards the null, as was shown empirically in a case–crossover study of the effects of long half-life benzodiazepines and the risk of motor vehicle crashes (MVC) in the elderly [40]. There were 5579 cases of MVC identified from the Province of Quebec, Canada, computerized databases. The case–crossover approach applied to all cases did not show any effect (OR 0.99; 95% CI 0.83–1.19). However, among the cases restricted to subjects with four or fewer prescriptions filled in the previous year (transient use), the odds ratio was 1.53 (95% CI 1.08–2.16]. Thus, it is important to verify this assumption of transient exposure, which may not be met in practice for drug therapies that are given for chronic conditions. This approach has been used successfully in several studies [41–45]. It has also been adapted for application to the risk assessment of vaccines (see Chapter 20) [46].

Case–Time–Control Design

One of the limitations of the case–crossover design is the assumption of the absence of a time trend in the exposure prevalence. An approach that adjusts for such time trends is the case–time–control method. By using cases and controls of a conventional case–control study as their own referents, the *case–time–control design* attempts to limit the biasing effect of unmeasured confounding factors, such as drug indication, while addressing the time trend assumption [47]. The method is an extension of the case–crossover analysis that uses, in addition to the case series, a series of control subjects to adjust for exposure time trends.

The approach is illustrated with data from the Saskatchewan Asthma Epidemiologic Project [14], a study conducted to investigate the risks associated with the use of inhaled beta-agonists in the treatment of asthma. Using a cohort of 12 301 asthmatics followed during 1980–87, 129 cases of fatal or near-fatal asthma and 655 controls were identified. The amount of beta-agonist used in the year prior to the index date was used for exposure. Table 43.2 displays the data comparing low (12 or fewer canisters per year) with high (more than 12) use of beta-agonists. The crude odds ratio for high beta-agonist use was 4.4 (95% CI 2.9–6.7). Adjustment for all available markers of severity, such as oral corticosteroids and prior asthma hospitalizations as confounding factors, lowers the odds ratio to 3.1 (95% CI 1.8–5.4), the "best" estimate one can derive from these case–control data using conventional tools.

To apply the case–time–control design, exposure to beta-agonists was obtained for the one-year current period and the one-year reference period prior to the current period. First, a

Table 43.2 Illustration of a case–time–control analysis of data from a case–control study of 129 cases of fatal or near-fatal asthma and 655 matched controls, and current beta-agonist use.

			Controls			
Cases	**High**	**Low**	**High**	**Low**	**OR**	**95% CI**
Current beta-agonist use (case–control)	93	36	241	414	3.1[b]	1.8–5.4
Discordant[a] use (case–crossover)	29	9			3.2	1.5–6.8
Discordant[a] use (control–crossover)			65	25	2.6	1.6–4.1
Case–time–control	29	9	65	25	1.2	0.5–3.0

[a] Discordant from exposure level during reference time period.
[b] Adjusted estimate from case–control analysis.
CI, confidence interval; OR, odds ratio.

case–crossover analysis was performed using the discordant subjects among the 129 cases, namely the 29 who were current high users of beta-agonists and low users in the reference period, and the nine cases who were current low users of beta-agonist and high users previously. This analysis is repeated for the 655 controls, of which there were 90 discordant in exposure; that is, 65 were current high users of beta-agonists and low users in the reference period, and 25 were current low users of beta-agonists and high users previously. The case–time–control odds ratio, using these discordant pairs frequencies for a paired-matched analysis, is given by (29/9)/(65/25) = 1.2 (95% CI 0.5–3.0). This estimate, which minimizes the effect of unmeasured confounding by disease severity, indicates a very small risk for these drugs.

The case–time–control approach can provide an unbiased estimate of the odds ratio in the presence of confounding by indication, despite the fact that the indication for drug use (in our example, intrinsic disease severity) is not measured, because of the within-subject analysis. It also controls for time trends in drug use. Nevertheless, its validity is subject to several assumptions, including the absence of time-dependent confounders, such as increasing asthma severity over time (an important problem, since new drugs may be more likely to be implemented when disease is most severe), so that caution is recommended in its use [48,49]. This approach has been used successfully in several studies [50–55].

Analytic Approaches for Improved Confounding Control

Balancing Patient Characteristics

Confounding caused by imbalance of patient risk factors between treatment groups is a known threat to validity in nonrandomized studies of treatment effects. A litany of options for reducing confounding is available to epidemiologists [56,57]. Several approaches fit key characteristics of longitudinal healthcare databases well and address important concerns in pharmacoepidemiologic analyses.

Propensity Score Analyses

Propensity score analysis has emerged as a convenient and effective tool for adjusting large numbers of confounders. In an incident user cohort design, a propensity score (PS) is the estimated probability of starting medication A versus starting medication B, conditional on all observed pretreatment patient characteristics. Such prediction of treatment choice based on preexisting patient characteristics fits the structure of the incident user cohort design.

Propensity scores are known as a multivariate balancing tool that balance large numbers of covariates in an efficient way even if the study outcome is rare, which is frequent in pharmacoepidemiology [58]. Estimating the propensity score using logistic regression is uncomplicated. Strategies for variable selection (i.e., the variables to include in the logistic regression model to estimate the propensity score) are well described [59]. Variables that are only predictors of treatment choice but are not independent predictors of the study outcome will lead to less precise estimates and in some extreme situations to bias [60]. Selecting variables based on *P* values is not helpful as this strategy depends on study size, and different variables would be selected or unselected for confounding adjustment if the study size changes, although the confounding effect of each variable may remain unchanged. Once a propensity score is estimated based on observed baseline covariates, there are several options to utilize it in a second step to adjust confounding. Typical strategies include adjustment for quintiles or deciles of the score with or without trimming, regression modeling of the PS, or matching on propensity scores [61]. Matching illustrates the working of propensity scores well.

Fixed ratio matching on propensity scores like 1:1 matching has several advantages that may outweigh its drawback of not utilizing the full dataset in situations where not all eligible patients match. Several matching algorithms are frequently used [62]. They have in common that for each exposed patient with a specific propensity score, one or multiple comparator patients will be picked with a propensity score that is similar within a defined caliper [63]. They vary in how they identify the best matches. Such matching will exclude patients in the extreme PS ranges where there is little clinical ambivalence in treatment choice; we therefore see little or no overlap in data (Figure 43.7). The tails of the PS distribution often harbor extreme patient scenarios caused by unobserved patient characteristics often in patients who are not representative for the majority in clinical practice. Keeping them in the analyses may lead to clinically less relevant

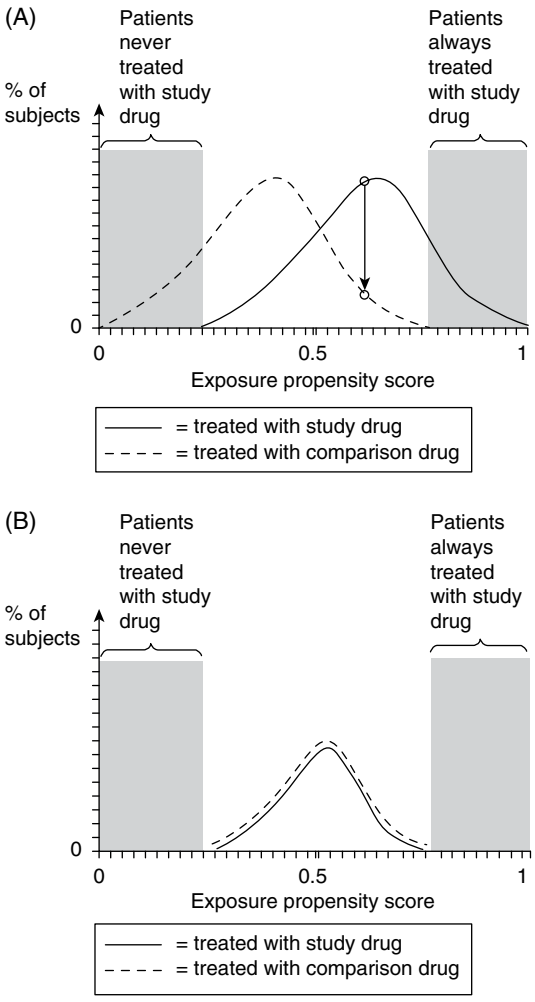


Figure 43.7 Two hypothetical propensity score distributions before and after matching. (A) Before matching: two propensity score distributions partially overlap, indicating some similarities between the comparison groups in a multivariate parameter space. (B) After 1:1 matching on propensity score: not all patients found matches that were similar enough in their multivariable characteristics. Areas of nonoverlap between PS distributions drop out entirely.

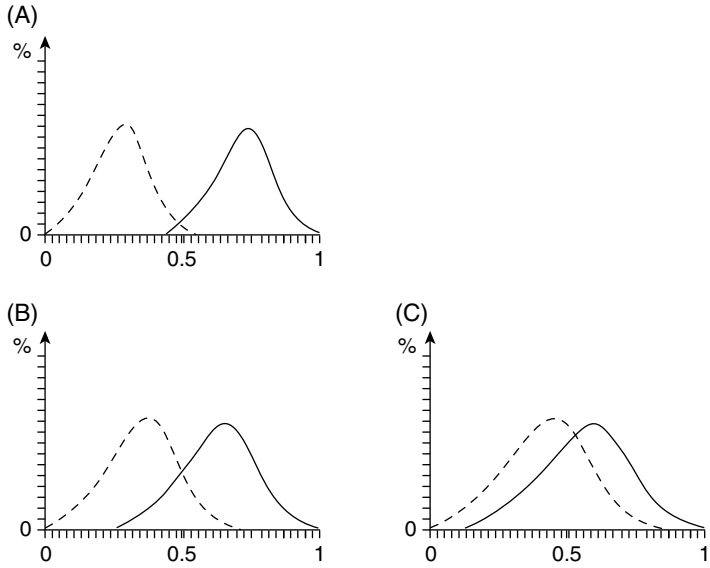


Figure 43.8 Multivariate propensity score distributions with varying degrees of overlap (A, low; B, moderate; C, high) as a diagnostic tool.

findings [64,65]. Trimming the extremes of the propensity score distributions is a data restriction strategy that generally will increase internal validity of findings [66]. Another advantage is that the multivariate balance of potential confounders can be demonstrated by cross-tabulating observed patient characteristics by actual exposure status after fixed ratio matching. 1:1 matching in cohort studies does not require matched analyses, which simplifies the effect estimation to a bivariate analysis. 1:1 matching allows inclusion of all overlapping comparator patients within a defined caliper in the analysis. However, in a variable ratio matching design the matching sets need to be preserved in the analysis to avoid bias. Analytic techniques that condition on the matching sets and may be used in this setting include conditional logistic regression or stratified Cox regression, depending on the data model.

It has been shown that on average, multivariate covariate balance will be achieved between treatment groups when matching on propensity score [67]. If a rational treatment decision process can be modeled well with observed patient characteristics, a resulting propensity score may lead to substantial or even full separation of treated and untreated patients (Figure 43.8A) [68]. This means that for patients initiated on a study drug, very few patients initiated on a comparison drug could be identified who had the same propensity for treatment given the observed patient characteristics. This would leave few comparable patients for analysis. In other words, treatment choice would be almost deterministic; little random treatment choice or empirical equipoise would be left in the prescribing decision that could be exploited for inference about the drug effect.

Consider the comparison of a fixed combination of ezetimibe and simvastatin versus simvastatin alone and their effect on coronary events as an example of such a situation. Assume that a health plan that provides the study data covers the ezetimibe/simvastatin combination only if LDL and HDL levels have crossed certain thresholds: every patient below those thresholds will use simvastatin alone. The LDL and HDL levels therefore become strong if not perfect determinants of treatment choice, and

including them in the propensity score estimation will lead to substantial or complete separation of the PS distributions of the two treatment groups. As the ezetimibe/simvastatin combination continues to be marketed, it will be used less selectively by more and more patients. Consequently, as the prescribers' treatment decisions are less disease state determined (e.g., not driven by LDL/HDL levels in the ezetimibe vs simvastatin example) and increasingly preference based, the propensity score distributions will overlap more and more as a sign that more patients are subject to treatment equipoise (Figure 43.8B,C).

If strong separation of PS distributions is observed, it indicates that the specific comparison cannot be made validly in the study population. In the above example, all ezetimibe and simvastatin users have high LDL level and hardly any simvastatin users have a comparable LDL level. Therefore, very few comparable patients are available for valid inference. This is not a limitation of the method, but rather a very insightful multivariate diagnostic describing the limitations inherent in a study population. The corresponding effect estimates from conventional multivariate outcome models will have substantial imprecision, reflecting the fact that few patients contribute to the estimation despite a large study size. Investigators may want to reconsider the comparison agent and choose a more comparable drug or use another study population where there is less treatment separation in clinical practice.

In summary, propensity score analyses are convenient tools to adjust for many covariates when study outcomes are rare. Extensive confounding adjustment is central in most pharmacoepidemiologic applications and in secondary healthcare databases we can often define many covariates in an effort to reduce the limitation of unobserved or misclassified patient characteristics. As such, PS analyses fit the needs of pharmacoepidemiologists working with longitudinal claims data well. In contrast to traditional outcome models, PS analyses allow the investigator to demonstrate the covariate balance achieved in the final study sample. Postmatching c-statistics or standardized differences of covariates have gained popularity in PS matching analyses [63,69]. PS estimation is well developed for comparing two agents using logistic regression to predict treatment choice. When more than two agents or several dose categories are compared, polytomous regression models are used to estimate the propensity score [70] and either pragmatic pairwise matching to a common reference group or multidimensional matching is applied [71]. Of importance, PS analyses adjust for measured variables, although they can be used to adjust for many at the same time some of which will be proxies for unobserved confounders [72]. Further, one loses the ability to see the effects of adjusting for one variable at a time.

In situations where exposure is rare, disease risk scores, an alternative to propensity score analysis, might be more suitable [73,74]. They estimate the association between patient factors and the study outcome in an unexposed population using multivariate regression and summarize the relationship in each patient's estimated probability of the outcome independent of exposure.

Focusing on the Analysis of Comparable Patients

Restriction is a common and effective analytic tool to make drug user groups more comparable by making populations more homogeneous, which leads to less residual confounding. Some restrictions are quite obvious since they are made by explicit criteria, for example, limiting the study population to elderly patients with dementia to study the safety of antipsychotic medications used to control behavioral disturbances in this population. Other restrictions, like PS matching, are more implicit and blur the line between design choices and analytic strategies to reduce confounding. It is important for pharmacoepidemiologists to understand the

reasons for specific restrictions and their implications for the generalizability of findings.

Choice of Comparator Group

Picking a comparator group is arguably the most fundamental choice in a pharmacoepidemiologic study design and may influence results substantially. Ideally, we want to restrict the comparison population to patients who have the identical indication as the users of the study agent in routine care. Rosiglitazone and pioglitazone are such a medication pair. They were marketed around the same time, were both indicated for second-line treatment of diabetes, come from the same class of compound, and in the early marketing phase were thought to have similar effectiveness and safety profiles. This should make treatment choice largely random with regard to patient characteristics and treatment groups comparable by design, resulting in almost overlapping propensity score distributions and little confounding (see Figure 43.8C). In individual situations, it may be that rosiglitazone-preferring physicians may treat less sick patients or independently produce better health outcomes in comparable patients. However, these physicians may or may not average out with similar pioglitazone-preferring physicians in this setting of treatment equipoise. As indications are usually recorded unreliably and frequently go beyond the labeled indications, picking a comparison drug that implicitly has the identical indication, if available, is usually more fruitful.

Limiting to Incident Users

By restricting the study population to new users of the study agent or a comparator agent, we implicitly require that both groups have been recently evaluated by a physician. Based on this evaluation, the physician has decided that the indicating condition has reached a state where a pharmacologic treatment should be initiated. Therefore, such patients are likely to be more similar in observable and unobservable characteristics than comparing incident users versus nonusers or versus ongoing users of another drug.

Matching on Patient Characteristics

Multivariate propensity scores demonstrate areas of nonoverlap where no referent patients with comparable baseline characteristics can be identified. It is recommended to remove those patients from the analysis as they are not contributing to the estimation and may introduce bias. Such a restriction can be achieved by trimming these patients from the study population [66] or by matching patients on the propensity score or on specific key patient characteristics of importance.

While restriction is an important tool to improve internal validity, it will reduce generalizability of findings. However, in pharmacoepidemiology we usually place higher value on internal validity even if that comes at the price of reduced external validity. Investigators will need to be aware of this trade-off and justify their choices accordingly.

Unobserved Patient Characteristics and Residual Confounding

Once a study is implemented, strategies to reduce confounding further are limited to observable disease risk factors. Secondary data, like electronic healthcare databases, often lack critical details on health state and risk factors, which leads to residual confounding if left unadjusted.

Proxy Adjustment

Longitudinal electronic healthcare databases are as much a description of medical sociology under financial constraints as they are records of delivered healthcare and can be analyzed as a set of proxies that indirectly describe the health status of patients [75]. This status is presented through the lenses of healthcare providers recording their findings and interventions with or without the help of professional coders and operating under the constraints of a specific healthcare system. On several steps along the

way, weighing of medical evidence and treatment options occurred; these are not observable in claims data but collectively resulted in a measurable action. A measured action like the filling of a medication has a clear interpretation but such interpretations are not always possible. In fact, in most cases we cannot determine the exact interpretation, but an exact interpretation may not be required for effective confounder adjustment. For example, old age serves as a proxy for many factors including co-morbidity, frailty, and cognitive decline; use of an oxygen canister is a sign of frail health; having regular annual check-ups is indicative of a health-seeking lifestyle and increased adherence. Adjusting for a perfect surrogate of an unmeasured factor is equivalent to adjusting for the factor itself [76].

The degree to which a surrogate is related to an unobserved or imperfectly observed confounder is proportional to the degree to which adjustment can be achieved [77,78]. Frequently used proxies in pharmacoepidemiologic analyses are the number of prescription drugs dispensed, the number of physician visits, and hospitalizations before the index drug exposure. Such measures of healthcare utilization intensity are useful proxies for general health, access to care, and surveillance. They have been shown to meaningfully help adjust for confounding [79].

Proxy adjustment can be exploited by algorithms that systematically search through recorded codes for diagnoses, procedures, equipment purchases, and drug dispensings to identify potential confounders or proxies thereof [72]. The hundreds of proxies that will be identified can then be adjusted for in a large propensity score model. Collinearity may likely occur in such large models. It will not affect estimation validity as the individual parameters estimated in the large propensity score regression will not be interpreted but only used for predicting treatment [58]; however, it may reduce precision [60]. This high-dimensional propensity score approach has been empirically shown to improve confounding adjustment in many settings over and above investigator-selected covariates [72,80–83].

While the semi-automated high-dimensional PS approach is remarkably robust, issues may arise in small studies with few exposed and rare outcomes [84,85]. Generally in such settings PS stratification is more robust [86] and variance estimates may be inflated [87]. Although adjusting for variables that are only related to the exposure and not to the outcome (an instrumental variable) could theoretically increase bias [60], in practical scenarios the advantage of adjusting for potential confounders outweighs the risk of adjusting for the rare instrument according to a recent simulation study [88]. A challenge remains that, empirically, it is very difficult to know with enough certainty whether a variable is a confounder or an instrument.

Exploiting Random Aspects in Treatment Choice Via Instrumental Variable Analysis

As explained earlier, we are interested in identifying residual random exposure variation after adjusting for observable confounders in order to more completely account for residual confounding. However, in secondary data such as longitudinal claims databases, electronic medical records, or registries, not all clinically relevant risk factors of the outcome may be recorded. To attempt to address this limitation, we can try to identify naturally occurring quasi-random treatment choices in routine care. Factors that determine such quasi-random treatment choices are called instrumental variables (IVs), and IV analyses can result in unbiased effect estimates even without observing all confounders if several assumptions are fulfilled (discussed further later).

An instructive example of an instrument is a hospital drug formulary. Some hospitals list only drug A for a given indication and other hospitals list only drug B. It is a reasonable assumption that patients do not choose their preferred hospital based on its formulary but rather on location and recommendation. Therefore, the

choice of drug A versus drug B should be independent of patient characteristics in the hospitals with these restricted formularies. Thus, comparing patient outcomes from drug A hospitals with patient outcomes from drug B hospitals should result in unbiased effects of drug A versus drug B, using the appropriate analytic tools. An example of such a study is one on the risk of death from aprotinin, an antifibrinolytic agent given to reduce bleeding during cardiac surgery [89]. The study identified surgeons who always used aprotinin and compared their outcomes to surgeons who always used aminocaproic acid, an alternative drug. If physician skill level and performance are on average equal between institutions, independent of drug use, this will result in valid findings. On the other hand, of course, such an assumption may not be valid, for example if academic hospitals allow less restrictive formularies, are more likely to see sicker patients, and have skilled physicians, all of which may be true.

Instrumental variable analyses rely on the identification of a valid instrument, a factor that is assumed to be related to treatment, but neither directly nor indirectly related to the study outcome. As such, an IV is an observed variable that causes (or is a marker of) variation in the exposure similar to random treatment choice. Typically, the following three assumptions need to be fulfilled for valid IV estimation: 1) an IV should affect treatment or be associated with treatment choice by sharing a common cause – the strength of this association is also referred to as the instrument strength; 2) an IV should be a factor that is as good as randomly assigned, so that it is unrelated to patient characteristics; and 3) an IV should not be related to the outcome other than through its association with treatment. As such, an IV analysis sounds very much like a randomized trial with noncompliance. The flip of a coin determines the instrument status (treat with A vs treat with B) and the amount of random noncompliance determines the strength of the instrument. In nonrandomized research, however, identifying valid instruments is difficult and successful IV analyses are infrequent. In principle, treatment preference can be influenced by time if treatment guidelines change rapidly and substantially. A comparison of patient outcome before versus after a sudden change in treatment patterns may then be a reasonable instrument [90,91]. Table 43.3 summarizes a list of some published IV analyses in healthcare.

More commonly, IV analyses utilize individual, local, or regional treatment preferences. For example, Brookhart *et al*. used physician prescribing preference to study the effect of analgesic treatment with COX-2 selective inhibitors (coxibs) versus nonselective NSAIDs (nsNSAID) on the risk of upper gastrointestinal (GI) bleed [93]. Many variations in defining this preference were tested and a reasonable instrument implementation turned out to be the same physician's analgesic prescription (IV status = coxib vs nsNSAID) to the previous patient who needed an analgesic [97]. The authors could demonstrate that such preference is a fairly strong instrument compared to instruments often used in economics. However, despite additional adjustment for observed patient characteristics and general quality of care [98,99], sicker patients may still cluster in coxib-preferring practices and be associated with GI bleed, which would invalidate the IV analysis. Stuckel et al. [94] used regional variation in the rate of cardiac catheterization (IV status = high vs low rate) to estimate its effect on post-MI mortality. While this regional preference instrument was weaker than the physician prescribing preference, it was argued that the instrument was more valid as it is less likely that patients would move to another region to receive the preferred care rather than simply switching their physician.

An IV analysis is technically fairly straightforward once all IV assumptions are fulfilled. In the case of a dichotomous instrument (Z) and exposure (X), the classic IV estimator is:

$$\beta_{IV} = \frac{E[Y \mid Z=1] - E[Y \mid Z=0]}{E[X \mid Z=1] - E[X \mid Z=0]}$$

Table 43.3 Selected examples of instrumental variable analyses in clinical epidemiology.

Instrument group	Instrument type	Examples
Sudden changes in treatment preference over time	Regulatory or coverage interventions	Johnston *et al.:* Beta-blocker use after heart failure hospitalization before and after 1998 [90]
	Innovations and rapid adoption	Juurlink *et al.:* Triamterene use in patients with hypertension before and after the RALES trial [91]
Provider treatment preference	Distance to specialist provider	McClellan *et al.:* Distance to cardiac cath lab facility in patients with acute myocardial infarction (MI) [92]
	Physician prescribing preference (PPP)	Brookhart *et al.:* Physician's treatment initiation choice to the preceding patient [93]
	Regional treatment preference	Stukel *et al.:* Variation of cardiac catheterization rates in 530 US regions in patients with MI [94]
	Hospital formulary/ surgeon treatment preference	Schneeweiss *et al.:* Cardiac surgeons who always use aprotinin as antifibrinolytic agent [89]
	Medication co-payment level	Cole *et al.:* Medication co-payment level in patients with heart failure and adherence [95]
	Dialysis center preference	Thamer *et al.:* Epo dosing by nonprofit vs for-profit dialysis centers [96]

where Y is the study outcome and β is a measure of the effect of X on Y [100]. The numerator of this estimator is the effect of the instrument status (coxib-preferring physician vs not) on the outcome measured as a risk difference. The denominator is the association between instrument status and actual treatment and is a measure of the strength of an instrument. In the case where the instrument perfectly predicts the treatment (e.g., in the example of a restrictive hospital formulary), then the denominator is 1 and the IV estimator will be identical to the naive risk difference estimate. As the instrument gets weaker, the denominator shrinks and the IV estimator increases relative to the naive risk difference estimate. The denominator is sometimes called a rescaling parameter as it scales up the naive risk difference estimate.

In practice, IV analyses use two-stage regression models that allow additional adjustment for multiple observed characteristics. These can be linear models to estimate risk differences or nonlinear models for risk ratio estimation [101]. Brookhart *et al*. have suggested several empirical tests to investigate the quality of an instrument in healthcare effectiveness research [4]. However, such strategies cannot test all assumptions and only help to rule out unsatisfactory IVs rather than confirm valid IVs. Fundamentally, the price of potentially unbiased estimation in IV analyses is the ultimately untestable assumptions that the authors will have to argue based on substantive knowledge and some empirical data. Because of the two-stage estimation, IV analyses are generally less precise which can, in some situations, severely reduce their utility for decision making. Users should also be cautioned that IV inference is based on those "marginal" patients whose treatment decision is influenced by the IV status. This concept is somewhat similar to propensity score analyses where only patients in the overlapping area of propensity score distributions contribute to the multivariate analysis. The IV analyses make an assumption of random treatment choice based on the nature of the healthcare system while

propensity score estimation is trying to utilize unexplained random treatment variation that is left after adjusting for all measured confounders.

Supplementing Database Studies with Clinically Rich Data on Potential Confounders

Resources and time permitting, another strategy to mitigate residual confounding is to identify a subsample and observe among a small number of patients detailed information on potential confounders (see sections earlier). A common version thereof is the nested case–control design or the case–cohort design where only a sample of controls or a sample of exposed and unexposed will be used to collect detailed confounder information. Eng *et al.* demonstrated the use of a case–cohort design embedded in a much larger claims-based analysis [102]. The two-stage sampling approach samples patients according to their exposure and outcome status simultaneously and then reweights findings [103]. Collet *et al.* demonstrated two-stage sampling in a Canadian healthcare database [104]. Increasingly, it is possible to link information-rich electronic health records or registry data in subsets of patients of large claims data studies. It is used to demonstrate that balance had been achieved in patient characteristics that were not observable in claims data [102]. In a new-user cohort study of oral antidiabetic medications with propensity score matching, it was demonstrated that laboratory test results, BMI, and duration of diabetes were well balanced although these parameters were only observable in the subset of EHR-linked patients and not part of the claims data analysis [105]. Such a process is less resource intensive and can be routinely applied in the right data environment.

From the perspective of secondary database studies, all these approaches can be described as internal validation studies, as patients are identified within the underlying study cohort and then contacted to retrieve more details on patient characteristics [106]. The advantage of these approaches is that they are tailored towards the specific question at hand, that is, the sampling as well as the confounder information of interest can be defined by the investigator. However, these approaches are operationally not necessarily efficient ways to collect information. They are often time-consuming since patients need to be identified and information needs to be collected.

An alternative approach is to utilize detailed confounder information that was already collected and then can be tied into the adjusted analysis of the main study cohort. If additional information is available elsewhere, such as a routinely conducted survey of a representative sample of the main database study, such external data sources can be used for reducing residual confounding under certain assumptions [107,108]. For example, each year the Medicare Current Beneficiary Survey routinely studies a representative sample of Medicare beneficiaries to measure a wide variety of characteristics that are not captured in Medicare claims data, for example limitations in activities of daily living [109], cognitive impairment, and physical impairments [110]. If such surveys are truly representative of the study cohort and data are already collected then such external adjustment has the advantage of being much faster and less costly. As the exact study question is not known when the external survey is conducted, it is recommendable to include a wide battery of patient characteristics in the questionnaire.

The available algebraic methods for such external adjustment [108] are limited to single binary confounders and cannot consider the joint confounding arising from several factors. These methods were recently extended to adjustment for multiple confounders of any scale using propensity score calibration (PSC) [111]. The basic concept of PSC is to estimate two multivariate propensity scores in the information-rich survey. One PS mimics the information available in the main study and is seen as an error-prone PS. The second PS uses all available information and is called the complete PS. By regressing the error-prone PS on the complete PS, a calibration factor can be estimated.

With this factor, the error-prone PS-adjusted result in the main study will be calibrated to produce results that are adjusted for the additional factors only available in the more detailed survey data using established regression calibration techniques [112]. Simulation studies have demonstrated good performance of PSC assuming that the relevant confounders were captured in the survey and the survey is representative of the main study [113]. PSC methods can be extended to other than survey data, including electronic medical records or disease registries.

Sensitivity Analyses

A series of sensitivity analyses can help investigators to better understand how robust a study's findings are to implicit and explicit assumptions. Some of the sensitivity analyses suggested below are generic and others are specific to database analyses.

An important but underutilized diagnostic tool for the impact of unobserved confounders on the validity of findings in nonrandomized studies is quantitative sensitivity analyses. Basic sensitivity analyses of residual confounding try to determine how strong and how imbalanced a confounder would have to be among drug categories to explain the observed effect. Such an "externally" adjusted relative risk (RR_{adj}) can be expressed as a function of the unadjusted relative risk (RR_{unadj}), the independent RR of the unmeasured confounder on the disease outcome (RR_{CD}), and the prevalence of the confounder in both drug exposure categories ($P_{C|E}$) [16]:

$$RR_{adj.} = \frac{RR_{unadj}}{\left[\frac{P_{C|E=1}\left(RR_{CD}-1\right)+1}{P_{C|E=0}\left(RR_{CD}-1\right)+1}\right]}$$

A recent cohort study could not find the expected association between use of TNF-alpha inhibitors, an immunomodulating agent, in treating rheumatoid arthritis, and the incidence of serious bacterial infections. There was a concern that physicians may have prescribed the agent selectively in patients with more progressive disease. A sensitivity analysis demonstrated the direction and strength of any such bias and concluded that it would be unlikely to change the clinical implications of the study [114]. This type of sensitivity analysis is particularly helpful in database studies, but is underutilized. Spreadsheet software is available for easy implementation of such sensitivity analyses (drugepi.org) [115]. Lash and Fink proposed an approach that considers several systematic errors simultaneously, allowing sensitivity analyses for confounding, misclassification, and selection bias in one process [116].

When using retrospective databases, it is usually cumbersome or impossible to contact patients and ask when they began using a drug for the first time in order to implement an incident user cohort design. Therefore, incident users are identified empirically by a drug dispensing that was not preceded by a dispensing of the same drug for a defined time period. This washout period is identical for all patients. A typical length is six months. In sensitivity analyses, this interval could be extended to nine and 12 months. In a study on the comparative safety of antidepressant agents in children in British Columbia, this interval was extended from one year to three years to ensure that the children in the study were treatment naive before their first use, which helped balance comparison groups and reduce confounding [117]. Although increasing the length of the washout increases the likelihood that patients are truly incident users, it also reduces the number of patients eligible for the study. This trade-off is particularly worth noting in health plans with high enrollee turnover.

There is often uncertainty about the correct definition of the exposure risk window based on the clinical pharmacology of the study agent. This is further complicated in healthcare databases, since the discontinuation date is imputed through the days' supply of the last dispensing/prescription. Varying the exposure risk window

is therefore insightful and easy to accomplish in cohort studies [118].

Another set of sensitivity analyses concerns the potential for informative censoring. Patients change and discontinue treatment because they lack a treatment effect or experience early signs of a side effect. The more strongly such nonadherence (i.e., drug switching or discontinuation) is associated with the outcome, the more an as-treated analysis, which censors at the point of discontinuation, will be biased. A cumulative risk analysis follows all patients for a fixed time period, carrying forward the initial exposure status and disregarding any changes in treatment status over time. Because this analysis disregards informative nonadherence, it will not suffer bias as a consequence of censoring, but it will suffer bias as a consequence of exposure misclassification. Such misclassification increases with a longer follow-up period and a shorter average time to discontinuation. In most cases, though not all, such misclassification will bias effects towards the null, similar to intention-to-treat analyses in randomized trials. Viewed separately, these two analysis types trade different biases but together, they give a range of plausible effect estimates. Adjusting for nonadherence in an analysis of a drug effect requires information about the predictors of treatment discontinuation [119,120], which is often not available with sufficient accuracy in pharmacoepidemiologic studies.

The Future

Minimizing confounding in nonrandomized pharmacoepidemiologic research is an ongoing development. While great progress has been made in analyzing longitudinal healthcare databases, much remains to be improved in order to reliably achieve unbiased estimates that will carry the weight of medical decision making. Several developments are promising. One is the use of instrumental variable analyses utilizing the multilevel structure of healthcare systems. Another is the expanded use of propensity score methods, including its combination with data-mining activities for high-dimensional proxy adjustment. A development that is gaining importance is the enrichment of existing data environments with supplemental clinical data linked from electronic medical records, disease registries, patient surveys, and/or laboratory test result repositories. While this information will provide an opportunity for improved confounding adjustment, it comes with equally large methodologic challenges, as information is collected in routine care and may have been requested/recorded selectively in patients who were thought to benefit most. Clearly, there is still plenty of work to be done to find satisfactory solutions for the control of confounding in the broad range of pharmacoepidemiologic research.

References

1 Walker AM. Confounding by indication. *Epidemiology* 1996; **7**(4): 335–6.

2 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**(9): 915–20.

3 Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010; **19**(8): 858–68.

4 Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 537–54.

5 Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009; **20**(4): 488–95.

6 Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008; **167**(4): 492–9.
7 Moride Y, Abenhaim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**: 731–7.
8 Schneeweiss S, Patrick AR, Sturmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–S142.
9 Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* 2010; **304**(8): 897–8.
10 Azoulay L, Assimes TL, Yin H, Bartels DB, Schiffrin EL, Suissa S. Long-term use of angiotensin receptor blockers and the risk of cancer. *PloS One* 2012; **7**(12): e50893.
11 Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; **29**(3): 479–86.
12 Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis appraisal by application to asbestos mining. *J Roy Stat Soc A* 1977; **140**: 469–91.
13 Lubin JH, Gail MH. Biased selection of controls for case–control analyses for cohort studies. *Biometrics* 1984; **40**: 63–75.
14 Spitzer WO, Suissa S, Ernst P, *et al*. The use of beta-agonists and the risk of death and near death from asthma. *N Engl J Med* 1992; **326**(8): 501–6.
15 Breslow N, Day NE. Statistical Methods in Cancer Research. Volume I – The Analysis of Case–Control Studies, 2nd edn. Lyon: International Agency for Research on Cancer, 1980.
16 Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Hagerstown: Lippincott-Raven, 1998.
17 Breslow N, Day NE. Statistical Methods in Cancer Research. Volume II – The Design and Analysis of Cohort Studies, 2nd edn. Lyon: International Agency for Research on Cancer, 1987.
18 Suissa S, Ernst P, Kezouh A, Bitton A, Hudson M. Newer disease modifying anti-rheumatic drugs (DMARDs) and the risk of serious hepatic adverse events in rheumatoid arthritis. *Am J Med* 2004; **117**(2): 87–92.
19 Suissa S, Hudson M, Ernst P. Leflunomide use and the risk of interstitial lung disease in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**(5): 1435–9.
20 Suissa S, Edwardes MD, Boivin JF. External comparisons from nested case–control designs. *Epidemiology* 1998; **9**(1): 72–8.
21 Rothman KJ. Modern Epidemiology, 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008.
22 Suissa S. The Quasi-cohort approach in pharmacoepidemiology: upgrading the nested case–control. *Epidemiology* 2015; **26**(2): 242–6.
23 Qian CJ, Coulombe J, Suissa S, Ernst P. Pneumonia risk in asthma patients using inhaled corticosteroids: a quasi-cohort study. *Br J Clin Pharmacol* 2017; **83**(9): 2077–86.
24 Anonymous. Risks of agranulocytosis and aplastic anemia. A first report of their relation to drug use with special reference to analgesics. The International Agranulocytosis and Aplastic Anemia Study. *JAMA* 1986; **256**(13): 1749–57.
25 Suissa S, Dell'aniello S, Martinez C. The multitime case–control design for time-varying exposures. *Epidemiology* 2010; **21**(6): 876–83.
26 Hutchinson GB. Leukemia in patients with cancer of the cervix uteri treated with radiation. *J Natl Cancer Inst* 1968; **40**: 951–82.
27 Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
28 Bergkvist L, Adami HO, Persson I, Hoover R, Schairer C. The risk of breast cancer after estrogen and estrogen–progestin replacement. *N Engl J Med* 1989; **321**: 293–27.

29 van der Klauw MM, Stricker BH, Herings RM, Cost WS, Valkenburg HA, Wilson JH. A population based case–cohort study of drug-induced anaphylaxis. *Br J Clin Pharmacol* 1993; **35**(4): 400–8.

30 Moulton LH, Wolff MC, Brenneman G, Santosham M. Case–cohort analysis of case–coverage studies of vaccine effectiveness. *Am J Epidemiol* 1995; **142**(9): 1000–6.

31 Strom BL, Schinnar R, Bilker WB, Feldman H, Farrar JT, Carson JL. Gastrointestinal tract bleeding associated with naproxen sodium vs ibuprofen. *Arch Intern Med* 1997; **157**(22): 2626–31.

32 Wacholder S. Practical considerations in choosing between the case–cohort and nested case–control designs. *Epidemiology* 1991; **2**: 155–8.

33 Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case–cohort designs. *J Clin Epidemiol* 1999; **52**(12): 1165–72.

34 Wacholder S, Boivin JF. External comparisons with the case–cohort design. *Am J Epidemiol* 1987; **126**(6): 1198–209.

35 Hemmelgarn B, Suissa S, Huang A, Boivin JF, Pinard G. Benzodiazepine use and the risk of motor vehicle crash in the elderly. *JAMA* 1997; **278**(1): 27–31.

36 Schneeweiss S, Doherty M, Zhu S, *et al.* Topical treatments with pimecrolimus, tacrolimus and medium- to high-potency corticosteroids, and risk of lymphoma. *Dermatology* 2009; **219**(1): 7–21.

37 Wacholder S. Practical considerations in choosing between the case–cohort and nested case–control designs. *Epidemiology* 1991; **2**(2): 155–8.

38 Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 459–68.

39 Maclure M. The case–crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991; **133**: 144–53.

40 Hebert C, Delaney JA, Hemmelgarn B, Levesque LE, Suissa S. Benzodiazepines and elderly drivers: a comparison of pharmacoepidemiological study designs. *Pharmacoepidemiol Drug Saf* 2007; **16**(8): 845–9.

41 Ray WA, Fought RL, Decker MD. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. *Am J Epidemiol* 1992; **136**: 873–83.

42 Sturkenboom MC, Middelbeek A, de Jong van den Berg LT, van den Berg PB, Stricker BH, Wesseling H. Vulvo-vaginal candidiasis associated with acitretin. *J Clin Epidemiol* 1995; **48**(8): 991–7.

43 Meier CR, Derby LE, Jick SS, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 1999; **281**(5): 427–31.

44 Barbone F, McMahon AD, Davey PG, *et al.* Association of road-traffic accidents with benzodiazepine use. *Lancet* 1998; **352**(9137): 1331–6.

45 Confavreux C, Suissa S, Saddier P, Bourdes V, Vukusic S. Vaccinations and the risk of relapse in multiple sclerosis. Vaccines in Multiple Sclerosis Study Group. *N Engl J Med* 2001; **344**(5): 319–26.

46 Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**(11): 1165–73.

47 Suissa S. The case–time–control design. *Epidemiology* 1995; **6**(3): 248–53.

48 Greenland S. Confounding and exposure trends in case–crossover and case–time–control design. *Epidemiology* 1996; **7**: 231–9.

49 Suissa S. The case–time–control design: further assumptions and conditions. *Epidemiology* 1998; **9**(4): 441–5.

50 Hernandez-Diaz S, Hernan MA, Meyer K, Werler MM, Mitchell AA. Case–crossover and case–time–control designs in birth

defects epidemiology. *Am J Epidemiol* 2003; **158**(4): 385–91.

51 Corrao G, Zambon A, Faini S, Bagnardi V, Leoni O, Suissa S. Short-acting inhaled beta-2-agonists increased the mortality from chronic obstructive pulmonary disease in observational designs. *J Clin Epidemiol* 2005; **58**(1): 92–7.

52 Kjaer D, Horvath-Puho E, Christensen J, *et al.* Use of phenytoin, phenobarbital, or diazepam during pregnancy and risk of congenital abnormalities: a case–time–control study. *Pharmacoepidemiol Drug Saf* 2007; **16**(2): 181–8.

53 Zambon A, Polo FH, Contiero P, Corrao G. Effect of macrolide and fluoroquinolone antibacterials on the risk of ventricular arrhythmia and cardiac arrest: an observational study in Italy using case–control, case–crossover and case–time–control designs. *Drug Saf* 2009; **32**(2): 159–67.

54 Park-Wyllie LY, Mamdani MM, Li P, Gill SS, Laupacis A, Juurlink DN. Cholinesterase inhibitors and hospitalization for bradycardia: a population-based study. *PLoS Med* 2009; **6**(9): e1000157.

55 Risselada R, Straatman H, van KF, *et al.* Platelet aggregation inhibitors, vitamin K antagonists, and risk of subarachnoid haemorrhage. *J Thromb Haemost* 2011; **9**(3): 517–23.

56 McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; **12**(7): 551–8.

57 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**(4): 323–7.

58 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**(8 Pt 2): 757–63.

59 Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010; **48**(6 Suppl): S114–S120.

60 Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006; **163**(12): 1149–56.

61 Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006; **98**(3): 253–9.

62 Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 2): 69–80.

63 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009; **28**(25): 3083–107.

64 Kurth T, Walker AM, Glynn RJ, *et al.* Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006; **163**(3): 262–70.

65 Lunt M, Solomon D, Rothman K, *et al.* Different methods of balancing covariates leading to different effect estimates in the presence of effect modification. *Am J Epidemiol* 2009; **169**(7): 909–17.

66 Sturmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution – a simulation study. *Am J Epidemiol* 2010; **172**(7): 843–54.

67 Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics*1985; **41**(1): 103–16.

68 Sekula P, Caputo A, Dunant A, *et al.* An application of propensity score methods to estimate the treatment effect of corticosteroids in patients with severe cutaneous adverse reactions.

Pharmacoepidemiol Drug Saf 2010; **19**(1): 10–18.

69 Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med* 2014; **33**(10): 1685–99.

70 Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Sturmer T. Confounder summary scores when comparing the effects of multiple drug exposures. *Pharmacoepidemiol Drug Saf* 2010; **19**(1): 2–9.

71 Solomon DH, Rassen JA, Glynn RJ, *et al*. The comparative safety of opioids for nonmalignant pain in older adults. *Arch Intern Med* 2010; **170**(22): 1979–86.

72 Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; **20**(4): 512–22.

73 Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res* 2009; **18**(1): 67–80.

74 Arbogast PG, Kaltenbach L, Ding H, Ray WA. Adjustment for multiple cardiovascular risk factors using a summary risk score. *Epidemiology* 2008; **19**(1): 30–7.

75 Schneeweiss S. Understanding secondary databases: a commentary on "Sources of bias for health state characteristics in secondary databases". *J Clin Epidemiol* 2007; **60**(7): 648–50.

76 Wolldridge JM. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press, 2001.

77 Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; **112**(4): 564–9.

78 Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985; **122**(3): 495–506.

79 Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol* 2001; **154**(9): 854–64.

80 Garbe E, Klos, S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. *Eur J Clin Pharmacol* 2013; **69**(3), 549–57.

81 Guertin JR, Rahme E, LeLorier J. Performance of the high-dimensional propensity score in adjusting for unmeasured confounders. *Eur J Clin Pharmacol* 2016; **72**(12): 1497–505.

82 Hallas J, Pottegard A. Performance of the high-dimensional propensity score in a Nordic healthcare model. *Basic Clin Pharmacol Toxicol* 2017; **120**(3), 312–17.

83 Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 41–9.

84 Patorno E, Glynn RJ, Hernandez-Diaz S, Liu J, Schneeweiss S. Studies with many covariates and few outcomes: selecting covariates and implementing propensity-score-based confounding adjustments. *Epidemiology* 2014; **25**(2): 268–78.

85 Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol* 2011; **173**(12): 1404–13.

86 Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF. A propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent. *Epidemiology* 2017; **28**(2): 249–57.

87 Schuster T, Lowe,WK, Platt RW. Propensity score model overfitting led to inflated variance of estimated odds ratios. *J Clin Epidemiol* 2016; **80**: 97–106.

88 Myers JA, Gagne JJ, Rassen JA, *et al*. [Abstract] Effects of adjusting for instrumental variables on bias and precision of effect estimates. International Conference for Pharmacoepidemiology, 2011.

89 Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008; **358**(8): 771–83.

90 Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 2008; **27**(9): 1539–56.

91 Juurlink DN, Mamdani MM, Lee DS, *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; **351**(6): 543–51.

92 McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994; **272**(11): 859–66.

93 Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**(3): 268–75.

94 Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA* 2007; **297**(3): 278–85.

95 Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: effect on cost and outcomes. *Pharmacotherapy* 2006; **26**(8): 1157–64.

96 Thamer M, Zhang Y, Kaufman J, Cotter D, Dong F, Hernan MA. Dialysis facility ownership and epoetin dosing in patients receiving hemodialysis. *JAMA* 2007; **297**(15): 1667–74.

97 Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application – in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; **62**(12): 1233–41.

98 Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis Rheum* 2006; **54**(11): 3390–8.

99 Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care* 2007; **45**(10 Suppl 2): S116–S122.

100 Angrist J.D, Imbens G.W, Rubin B. Identification of Casual effects Using Instrumental Variables. Journal of the American Statistical Association 2002.

101 Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; **169**(3): 273–84.

102 Eng PM, Seeger JD, Loughlin J, Clifford CR, Mentor S, Walker AM. Supplementary data collection with case–cohort analysis to address potential confounding in a cohort study of thromboembolism in oral contraceptive initiators matched on claims-based propensity scores. *Pharmacoepidemiol Drug Saf* 2008; **17**(3): 297–305.

103 Walker AM. Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. *Biometrics* 1982; **38**(4): 1025–32.

104 Collet JP, Schaubel D, Hanley J, Sharpe C, boivin JF. Controlling confounding when studying large pharmacoepidemiologic databases: a case study of the two-stage sampling design. *Epidemiology* 1998; **9**(3): 309–15.

105 Patorno E, Gopalakrishnan C, Franklin JM, *et al.* Claims-based studies of oral glucose-lowering medications can achieve balance in critical clinical parameters only observed in electronic health records. *Diabetes Obes Metab* 2018; **20**(4): 974–84.

106 Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med Care* 2007; **45**(10 Supl 2): S158–S165.

107 Suissa S, Edwardes MD. Adjusted odds ratios for case–control studies with missing confounder data in controls. *Epidemiology* 1997; **8**(3): 275–80.

108 Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. *Epidemiology* 2005; **16**(1): 17–24.

109 Saliba D, Orlando M, Wenger NS, Hays RD, Rubenstein LZ. Identifying a short functional disability screen for older persons. *J Gerontol A Biol Sci Med Sci* 2000; **55**(12): M750–M756.

110 Eppig FJ, Chulis GS. Matching MCBS (Medicare Current Beneficiary Survey) and Medicare data: the best of both worlds. *Health Care Financ Rev* 1997; **18**(3): 211–29.

111 Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol* 2005; **162**(3): 279–89.

112 Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990; **132**(4): 734–45.

113 Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Propensity score calibration and its alternatives. *Am J Epidemiol* 2007; **165**(10): 1122–3.

114 Psaty BM, Koepsell TD, Lin D, *et al*. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999; **47**(6): 749–54.

115 Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006; **15**(5): 291–303.

116 Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**(4): 451–8.

117 Schneeweiss S, Patrick AR, Solomon DH, *et al*. Comparative safety of antidepressant agents for children and adolescents regarding suicidal acts. *Pediatrics* 2010; **125**(5): 876–88.

118 McMahon AD, Evans JM, McGilchrist MM, McDevitt DG, MacDonald TM. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 1998; **7**(4): 275–80.

119 Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**(3): 779–88.

120 Toh S, Hernandez-Diaz S, Logan R, Rossouw JE, Hernan MA. Coronary heart disease in postmenopausal recipients of estrogen plus progestin therapy: does the increased risk ever disappear? A randomized trial. *Ann Intern Med* 2010; **152**(4): 211–17.

Part VI

Conclusion

44

The Future of Pharmacoepidemiology

Brian L. Strom[1], Stephen E. Kimmel[2], and Sean Hennessy[2]

[1] *Rutgers Biomedical and Health Sciences, Newark, NJ, USA*
[2] *Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*

> *We should all be concerned about the future because we will have to spend the rest of our lives there.*
>
> (Charles Franklin Kettering, 1949)

Speculating about the future is at least risky and possibly foolish. Nevertheless, the future of pharmacoepidemiology seems apparent in many ways, judging from past trends and recent events. Interest in the field by the pharmaceutical industry, government agencies, new trainees, and the public continues to grow, as does realization of what pharmacoepidemiology can contribute. Indeed, international attention on drug safety remains high, important safety questions involving widely used drugs continue to emerge, and questions concerning the effectiveness of systems of drug approval and drug safety monitoring remain.

As the functions of academia, industry, and government have become increasingly global, so has the field of pharmacoepidemiology. The number of individuals attending the annual International Conference on Pharmacoepidemiology has increased from approximately 50 in the early 1980s to over 1800 in 2019. The International Society for Pharmacoepidemiology (ISPE), established in 1991, has grown to approximately 1500 members from over 60 countries. It developed a set of guidelines for Good Epidemiologic Practices for Drug, Device, and Vaccine Research in the United States in 1996 [1], and updated these guidelines most recently in 2016 as the ISPE Guidelines for Good Pharmacoepidemiology Practices [2]. Many national pharmacoepidemiologic societies have been formed as well. The journal *Clinical Pharmacology and Therapeutics*, the major US academic clinical pharmacology journal, actively solicits pharmacoepidemiologic manuscripts, as did the *Journal of Clinical Epidemiology*. The major journal devoted to the field, *Pharmacoepidemiology and Drug Safety*, ISPE's official journal, is indexed on Medline and achieved an impact factor of 2.314 in 2018, remarkably high for a niche field. Other journals have been formed to copy and compete with it.

The number of individuals seeking to enter the field continues to increase, as does their level of training. The number of programs of study in pharmacoepidemiology is increasing in schools of medicine, public health, and pharmacy. While in the 1980s the single summer short course in pharmacoepidemiology at the University of Minnesota was sometimes cancelled because of

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

insufficient interest, later the University of Michigan School of Public Health summer course in pharmacoepidemiology attracted 10% of all students in the entire summer program, and now McGill University, Erasmus University Rotterdam, Utrecht University, and the Johns Hopkins Bloomberg School of Public Health all conduct summer short courses in pharmacoepidemiology. Several other short courses are given as well, including by ISPE itself which has seen a massive increase in preconference courses offered over the years. Regulatory bodies around the world have expanded their internal pharmacoepidemiologic programs. The number of pharmaceutical companies with their own pharmacoepidemiologic units has also increased, along with their support for academic units and their funding of external pharmacoepidemiologic studies. Requirements that a drug be shown to be cost-effective (see Chapter 34) have been added to many national and provincial healthcare systems, and managed care organizations, either to justify reimbursement or even to justify drug availability. Drug utilization review is being widely applied (see Chapter 19), and many hospitals are becoming mini-pharmacoepidemiologic practice and research laboratories.

The US Congress has recognized the importance of pharmacoepidemiology, requiring the FDA to build a new data resource, containing at least 100 million lives, for evaluating potential adverse effects of medical products (see Chapter 25), and most recently passing the 21st Century Cures Act, encouraging the wide use of "real-world evidence." The latter has been deemed to range from traditional pharmacoepidemiology data sources like claims and medical record databases, and even *ad hoc* pharmacoepidemiology studies (see Part III), to novel data sources like e-health tools, m-health tools, and other wearable devices, as well as pragmatic trials using traditional pharmacoepidemiology databases to collect outcomes (see Chapter 32). The future is likely to see a marked expansion of these novel, technology-driven approaches.

Thus, from the perspective of those in the field, the trends in pharmacoepidemiology are remarkably positive, although many important challenges remain. In this chapter, we will briefly give our own view on the future of pharmacoepidemiology. Following the format of Part II of the book, we explore this future from the perspectives of academia, the pharmaceutical industry, regulatory agencies, and then the law.

The View from Academia

Scientific Developments

Methodologic Advances

Methodologically, the array of approaches available for performing pharmacoepidemiologic studies will continue to grow. Each of the methodologic issues discussed in Part V can be expected to be the subject of further research and development. The future is likely to see ever more advanced ways of performing and analyzing epidemiologic studies across all content areas, as the field of epidemiology continues to expand and develop. Some of these new techniques will, of course, be particularly useful to investigators in pharmacoepidemiology (see, for example, Chapter 43). The next few years will likely see continued expanded use of propensity scores, instrumental variables, the trend-in-trend design, sensitivity analysis, and novel methods to analyze time-varying exposures and confounders. In addition, we believe that we will see increasing application of pharmacoepidemiologic insight in the conduct of clinical trials, as well as increased use of the randomized trial design to examine questions traditionally addressed by observational pharmacoepidemiology (see Chapter 32), especially given the controversies resulting from inconsistencies between nonexperimental studies and experimental studies, and given the emerging field of comparative effectiveness research (see Chapter 26).

Drug regulators have enthusiastically embraced therapeutic risk management (see Chapter 24). Yet, this field is still very much in its infancy, with an enormous amount of work needed to develop new methods to measure, communicate, and manage the risks and benefits associated with medication use. Rigorous studies (i.e., program evaluations) of the effectiveness of risk management programs remain the exception rather than the rule. Development of this area will require considerable effort from pharmacoepidemiologists as well as those from other fields.

We may see developments in the processes used to assess causality from individual case reports (see Chapters 10 and 29). Data mining approaches will be used increasingly in spontaneous reporting databases to search for early signals of adverse reactions (see Chapter 27). Hopefully, we will see studies evaluating the utility of such approaches. The need for newer methods to screen for potential adverse drug effects, such as those using healthcare claims or medical record data and data from social media, is also clear.

We are likely to see increasing input from pharmacoepidemiologists into policy questions about drug approval (see Chapter 8), with new attention to applying pharmacoepidemiology in the study of the growing opiate epidemic (see Chapter 28). We anticipate that emphasis will shift from studies evaluating whether a given drug is associated with an increased risk of a given event to those that also examine patient- and regimen-specific factors that affect risk [3] (see also Chapters 24, 26, and 42). Such studies are crucial because, if risk factors for adverse reactions can be better understood before a safety crisis occurs, or early in the course of a crisis, then the clinical use of the drug may be able to be repositioned, avoiding the loss of useful drugs (see Chapters 24, 30, 35, and 39).

With recent developments in molecular biology and bioinformatics, and their application to the study of pharmacogenetics, the ability of researchers to identify biologic factors that predispose patients to adverse drug reactions has increased [4] (see Chapter 30). However, few of these discoveries have yet been shown useful in improving patient care, and new studies and methods must be pursued to determine the clinical utility of genetic testing. Pharmacogenetics has evolved from studies of measures of slow drug metabolism as a contributor to adverse reactions [5] to the study of molecular genetic markers [6–9]. This has been aided by the development of new, noninvasive methods to collect and analyze biosamples, making population-based genetic studies feasible. We believe that clinical measurement of biologic factors will ultimately complement existing approaches to tailoring therapeutic approaches for individual patients. However, it is unlikely that genotype will be the only, or even the major, factor that determines the optimal drug or dose for a given patient.

Future years are likely to see much more of this cross-fertilization between pharmacoepidemiology and molecular biology, and newer forms of "-omics" such as the microbiome. From a research perspective, we can easily envision pharmacogenetic studies added to the process of evaluating potential adverse reactions. We also anticipate the availability of genotypic information for members of large patient cohorts for whom drug exposures and clinical outcomes are recorded electronically, and even for selected patients from electronic data systems, such as those described in Part IIIb of this book.

New Content Areas of Interest

In addition, there are a number of new content areas that are likely to be explored more. Studies of drug utilization will continue to become more innovative (see Chapter 18). Particularly as the healthcare industry becomes more sensitive to the possibility of overutilization, underutilization, and inappropriate utilization of drugs, and the risks associated with each, one would expect

to see an increased frequency of and sophistication in drug utilization review programs, which seek to improve care (see Chapter 19), potentially incorporating techniques from molecular pharmacoepidemiology (see Chapter 30).

The US Joint Commission on Accreditation of Healthcare Organizations revolutionized US hospital pharmacoepidemiology through its standards requiring adverse drug reaction surveillance and drug use evaluation program in every hospital [10,11]. Hospitals are also now experimenting with different methods of organizing their drug delivery systems to improve their use of drugs, for example through use of computerized clinical decision support and the addition of pharmacists to patient care teams [12] (see Chapter 41).

Interest in the field of pharmacoeconomics, that is, application of the principles of health economics to the study of drug effects, is continuing (see Chapter 34). Society is realizing that the acquisition cost of drugs is often a very minor part of their economic impact, and that their beneficial and harmful effects can be vastly more important. Further, more governments and insurance programs are increasingly requiring economic justification before permitting reimbursement for a drug. As a result, the number of studies exploring this is increasing. As the methods of pharmacoeconomics become increasingly sophisticated, and its applications clear, this could be expected to continue to be a popular field of inquiry.

More nonexperimental studies of beneficial drug effects, particularly of drug effectiveness, can be expected, as the field becomes more aware that such studies are possible (see Chapter 33). This is being encouraged by the rapid increase in the use of propensity scores to adjust for measured covariates, although investigators using this method often place more confidence in that technique than is warranted, some not recognizing that its ability to control for confounding by indication remains dependent on one's ability to *measure* the true determinants of exposure (see Chapter 43). It is also being encouraged by the development of comparative effectiveness research (see Chapter 26). Other approaches to controlling for confounding are similarly likely to become more common as they are further developed (see Chapter 43). New analytic approaches, like machine learning, artificial intelligence, and cognitive computing, are also likely to make their way into pharmacoepidemiology studies.

We will also see more use of pharmacoepidemiologic approaches prior to drug approval, for example to understand the baseline rate of adverse events that one can expect to see in patients who will eventually be treated with a new drug (see Chapter 7).

Recent years have seen an explosion in the worldwide use of herbal and other complementary and alternative medications. These are essentially pharmaceuticals sold without conventional standardization, and with no required premarketing testing of safety or efficacy. In a sense, for these products, this is a return to a preregulatory era. Therefore, it is quite likely that the next few years will see an analogous set of safety concerns associated with their use, and society will turn to pharmacoepidemiologists to help evaluate the use and effects of these products. Of course, if regulatory oversight is decreased in some countries, as has been suggested in the US, the same could occur with traditional medications.

Research interest in the entire topic of patient nonadherence with prescribed drug regimens goes back to about 1960, but little fruitful research could be done because methods for ascertaining drug exposure in individual ambulatory patients were grossly unsatisfactory [13]. This problem has been mitigated greatly by advances in incorporating time-stamping microcircuitry into pharmaceutical containers, which records the date and time whenever the container is opened [14]. Perhaps as a consequence of its inherent simplicity and economy, electronic monitoring is increasingly emerging

as the *de facto* gold standard for compiling dosing histories of ambulatory patients, from which one can evaluate the extent of adherence to the prescribed drug regimen. Future years are likely to see a continuing increase in the use of this technique (see Chapter 38) in research, and perhaps in clinical practice. Perhaps equally importantly, new methods of measuring adherence that do not rely on purchasing and using alternative sources of drug dispensing, such as smartphone-based measures of pill taking, may expand our ability to measure adherence in real-world epidemiology studies.

The next few years are also likely to see the increasing ability to target drug therapy to the proper patients. This will involve increasing use of both statistical methods and laboratory techniques from other biological sciences, as described above. Statistical approaches will allow us to use predictive modeling to study, from a population perspective, who is most likely to derive benefit from a drug, and who is at greatest risk of an adverse outcome. Laboratory science will enable us to measure individuals' genotypes, to predict responses to drug therapy (i.e., molecular susceptibility). From the perspective of preapproval testing, these developments may allow researchers to target specific patient types for enrollment into their studies, those subjects most likely to succeed with a drug. From a clinical perspective, it will enable healthcare providers to incorporate biological factors in the individualization of choice of regimens.

The past few years have seen the increased use of surrogate markers, presumed to represent greater risk of rarer serious adverse effects when drugs are used in broader numbers of patients. These range from mild liver function test abnormalities, used as predictors of serious liver toxicity, to electrocardiographic QTc prolongation as a marker of risk of suffering the arrhythmia torsades de pointes, which can lead to death. Indeed, some drugs have been removed from the market, or from development, because of the presence of these surrogate markers. Yet the utility of these markers as predictors of serious clinical outcomes is poorly studied. The next few years are likely to see the increased use of both very large observational studies and large simple trials after marketing, to study important clinical outcomes (see Chapters 32 and 36).

In addition, with the growth of concerns about patient safety (see Chapter 41), there has been more attention to simultaneous use of pairs of drugs that have been shown in pharmacokinetic studies (see Chapter 2) to cause increased or decreased drug levels. Yet population studies informing the clinical importance and pharmacologic aspects of drug–drug interactions have only been performed in the past few years (see Chapter 40). The next few years are likely to see the emergence of more studies to address such questions.

Finally, in the last few years, society has increasingly turned to pharmacoepidemiology for input into major policy decisions. For example, pharmacoepidemiology played a major role in the evaluations by the Institute of Medicine of the US National Academy of Sciences of the anthrax vaccine [15] (deciding whether the existing vaccine was safe to use and, thereby, whether the military vaccine program should be restarted) and the smallpox vaccine program (deciding the shape of the program intended initially to vaccinate the entire US population) [16]. This is likely to occur even more often in the future.

Logistical Advances

Logistically, with the increased computerization of data in society in general and within healthcare in particular, and the increased emphasis on using electronic databases for pharmacoepidemiology [17] (see Part IIIb), some data resources will disappear (e.g., the Rhode Island Drug Use Reporting System and the inpatient databases discussed in prior editions of this book have disappeared, with new ones added,

and Group Health of Puget Sound has become less commonly used as a data resource, as much larger databases have emerged), and a number of new computerized databases have emerged as major resources for pharmacoepidemiologic research, such as commercial insurance databases (see Chapter 12), inpatient databases (see Chapter 14), and the databases from Ontario and Denmark (see Chapter 12). The importance of these databases to pharmacoepidemiology is now clear: they enable researchers to address, quickly and relatively inexpensively, questions about drug effects in different settings that require large sample sizes, with excellent quality data on drug exposures. Registries (see Chapter 16) will also become increasingly important for pharmacoepidemiologic research. With the initiation of US Medicare Part D in 2006, which provides prescription drug coverage to US Medicare recipients, the availability of this data resource is potentially "game changing" for hypothesis testing studies, as it is so large relative to other resources; nearly 27 million Medicare beneficiaries were already subscribed to Part D coverage in 2009 [18] (see Chapter 12). It has created an enormous new data resource for pharmacoepidemiology, as well as increased interest from the US government in what pharmacoepidemiology can do. The development of the FDA's Sentinel Initiative [19] (see Chapter 25) has similarly provided a vast new data resource, initially intended for hypothesis generating, and more recently used for hypothesis strengthening and testing.

Nevertheless, even as the use of databases increases, it is important to keep in mind the importance of studies that collect data *de novo* (see Chapter 16). Each approach to pharmacoepidemiology has its advantages and its disadvantages, as described in Part III. No approach is ideal in all circumstances, and often a number of complementary approaches are needed to answer any given research question (see Chapter 17). To address some of the problems inherent in any database, we must maintain the ability to perform *ad hoc* studies as well (see Chapter 16). Perhaps better, less expensive, and complementary approaches to *ad hoc* data collection in pharmacoepidemiology will be developed. For example, a potential approach that has not been widely used is the network of regional and national poison control centers. In particular, poison control centers would be expected to be a useful source of information about dose-dependent adverse drug effects.

Of critical importance, there is increasing concern about patient privacy in many countries. The regulatory framework for human research is actively changing in the process, such as Europe's new data protection law. As discussed in Chapter 31, this is already beginning to make pharmacoepidemiologic research more difficult, whether it affects access to medical records in database studies or access to a list of possible cases with a disease to enroll in *ad hoc* case–control studies. This will be an area of great interest and rapid activity over the next few years as electronic health records become much more commonplace, and one in which the field of pharmacoepidemiology will need to remain very active or risk considerable interference with its activities.

It is likely that new types of research opportunities will emerge. For example, as the US finally implemented a drug benefit as part of Medicare, its health program for the elderly, US government drug expenditures suddenly increased by $49.5 billion in 2007 [20]. Outside the US, many different opportunities to form databases are being developed. There is also increased interest in the importance of pharmacoepidemiology in the developing world. Many developing world countries spend a disproportionate amount of their healthcare resources on drugs [21], yet these drugs are often used inappropriately [22]. There have been a number of initiatives in response to this, including the World Health Organization's development of its list of Essential Drugs [23,24].

Funding

For a number of years, academic pharmacoepidemiology suffered from limited research funding opportunities. In the early 1980s, the only available US funding for the field was an extramural funding program from the FDA with a total of $1 million/year. Industry interest and support were similarly limited. With growing interest in the field, this situation appears to be changing rapidly. The FDA has expanded its internal pharmacoepidemiology program and the US National Institutes of Health (NIH) is funding pharmacoepidemiologic studies. In the US, other funding now comes from the Agency for Health Care Research and Quality (AHRQ), and from the Patient-Centered Outcomes Research Institute (PCORI), created as part of the Affordable Care Act. Much industry funding is available, as the perceived need for the field within industry grows (see below). This is likely to increase, especially as the FDA more often requires industry to perform postmarketing studies, and with the legislative mandate for the FDA to pay more attention to "real-world evidence."

There is, of course, a risk associated with academic groups becoming too dependent on industry funding, in terms of both choice of study questions and credibility. Fortunately, in the US, the AHRQ has begun to fund pharmacoepidemiologic research as well, as part of an initiative in pharmaceutical outcomes research. In particular, the AHRQ Centers for Education and Research on Therapeutics (CERTs) program provided federal support for ongoing pharmacoepidemiologic activities (see Chapter 6). While still small relative to industry expenditures on research, it was large relative to the US federal funding previously available for pharmacoepidemiology. Similar programs have now been started in Europe and Canada. Unfortunately, the CERTs program has ended and the future of the AHRQ itself is always in question.

Even the US NIH now funds pharmacoepidemiologic projects more often. The NIH is the logical major source for such support, as it is the major funding source for most basic biomedical research in the US. Its funds are also accessible to investigators outside the US, via the same application procedures. However, the NIH's current organizational structure represents an obstacle to pharmacoepidemiologic support. In general, the institutes within the NIH are organized by organ system.

Earlier in the development of pharmacoepidemiology, the National Institute of General Medical Sciences (NIGMS) provided most of the US government support for our field. It remains, conceptually, perhaps the most appropriate source of such support, since it is intended to fund projects that are not specific to an organ system, and it is the institute that funds clinical pharmacologic research. However, over the past few years there has been limited funding from the NIGMS for epidemiologic research. A notable exception was the NIGMS-funded Pharmacogenetics Research Network (PGRN), which has now been disbanded. In the meantime, NIH funding continues to be available if one tailors a project to fit an organ system or in some other way fits the priorities of one of the individual institutes, such as the National Institute on Aging or the National Institute of Child Health and Human Development.

Personnel

With the major increase in interest in the field of pharmacoepidemiology, accompanied by an increased number of funding opportunities, a major remaining problem, aggravated by the other trends, is one of inadequate personnel resources. There is a desperate need for more well-trained people in the field, with employment opportunities available in academia, industry, and government agencies. Some early attempts were made to address this. The Burroughs Wellcome Foundation created the Burroughs Wellcome

Scholar Award in Pharmacoepidemiology, a faculty development award designed to bring new people into the field. This program did not provide an opportunity for fellowship training of entry-level individuals but was designed for more experienced investigators. Unfortunately, it is no longer active.

Outside government, training opportunities are limited. In the US, the NIH is the major source of support for scientific training but as noted above, the NIGMS, which funds training programs in clinical pharmacology, now supports one program in pharmacoepidemiology, while the National Heart, Lung and Blood Institute supports another. The National Institute of Child Health and Human Development also has funded limited training in pediatric pharmacoepidemiology. However, pharmacoepidemiologic training is still too dependent on nonfederal sources of funds, especially at a time when such funding is becoming harder to obtain.

There is a growing number of institutions now capable of carrying out such training, for example universities with faculty members interested in pharmacoepidemiology, including those with clinical research training programs supported by, for example, an NIH Clinical and Translational Science Award and organ system-specific training grants. Young scientists interested in undergoing training in pharmacoepidemiology, however, can only do so if they happen to qualify for support from such programs. No ongoing support is normally available from these programs for training in pharmacoepidemiology *per se*. This was addressed in the past primarily through the leadership and generosity of some pharmaceutical companies. Much more is needed, however. Fortunately, with the rapid rise in interest in comparative effectiveness research (see Chapter 26), additional training support emerged from both the NIH and AHRQ/PCORI, but this is now in question going forward. Further, the focus on comparative effectiveness research (see Chapter 26) and patient engagement and patient-reported outcomes (see Chapter 42) triggered by the PCORI are in doubt now, as PCORI was created by the Affordable Care Act which is itself at risk.

The View from Industry

It appears that the role of pharmacoepidemiology in industry is expanding rapidly. All that was said above about the future of pharmacoepidemiology scientifically, as it relates to academia (see Chapter 6), obviously relates to industry as well (see Chapter 7). The necessity of pharmacoepidemiology has become apparent to many of those in industry. In addition to being useful for exploring the effects of their drugs, manufacturers are beginning to realize that the field can contribute not only to identifying problems but also to documenting drug safety and developing and evaluating risk management programs. An increasing number of manufacturers are mounting pharmacoepidemiologic studies "prophylactically," to have safety data available in advance of when crises may occur. Proper practice would argue for postmarketing studies for all newly marketed drugs used for chronic diseases, and all drugs expected to be either pharmacologically novel or sales blockbusters, because of the unique risks that these situations present. Pharmacoepidemiology also can be used for measuring beneficial drug effects (see Chapter 33) and even for marketing purposes, in the form of descriptive market research and analyses of the effects of marketing efforts.

Perhaps most importantly for the industry's financial bottom line, pharmacoepidemiologic studies can be used to protect the major investment made in developing a new drug against false allegations of adverse effects, protecting good drugs for a public that needs them. Further, even if a drug is found to have a safety problem, the legal liability of the company may be diminished if the company has, from the outset, been forthright in its efforts to learn about that drug's

risks. Finally, as noted in Chapter 1 and above, the FDA now has new authority to require postmarketing pharmacoepidemiologic studies, and a new charge to focus on "real-world evidence," so one can expect to see much more required of industry by regulators.

Industry is always interested in predictability. With that, there is increased interest in developing a formulaic approach to risk–benefit assessment (see Chapter 35). The next few years are likely to see considerable additional work in this area.

In light of these advantages, most major pharmaceutical firms have formed their own pharmacoepidemiologic units. Of course, this then means that industry confronts and, in fact, aggravates the problem of an insufficient number of well-trained personnel. Many pharmaceutical companies increased their investment in external pharmacoepidemiologic data resources, so that they will be available for research when crises arise. This has been declining, however. A risk of the growth in the number of pharmacoepidemiologic studies for industry is the generation of an increased number of false signals about harmful drug effects. This is best addressed by having adequately trained individuals in the field, and by having personnel and data resources available to address these questions quickly, responsibly, and effectively, when they are raised.

The View from Regulatory Agencies

It appears that the role of pharmacoepidemiology in regulatory agencies is also expanding (see Chapter 8). Again, all of what was said above about the future of pharmacoepidemiology scientifically, as it relates to academia, obviously relates to regulatory agencies as well. In addition, there have been a large number of major drug crises, many described throughout this book. Many of these crises resulted in the removal of the drugs from the market. The need for and importance of pharmacoepidemiologic studies have become clear. Again, this can be expected to continue in the future. It has even been suggested that postmarketing pharmacoepidemiologic studies might replace some premarketing Phase III studies in selected situations, as was done with zidovudine [25]. As noted, regulatory agencies are being given increased authority to require such studies after marketing. They are also expanding their pharmacoepidemiologic staffing, and seeking training in pharmacoepidemiology for those already employed by the agencies.

We are also seeing increasing governmental activity and interest in pharmacoepidemiology, outside the traditional realm of regulatory bodies. For example, in the US, pharmacoepidemiology used to play an important role within the AHRQ, the Centers for Disease Control and Prevention, PCORI, and the NIH, and there has been for nearly 40 years intermittent debate about the wisdom of developing an independent new Center for Drug Surveillance [26–29].

As noted above, the use of therapeutic risk management approaches (see Chapter 24) has been aggressively embraced by regulatory bodies around the world, and there has been considerable discussion about risk–benefit assessments of medical products (see Chapter 35). This will continue to change regulation as more experience with it is gained.

There is considerable regulatory interest in getting important new drugs onto the market quickly, using mechanisms such as the FDA's initiatives on orphan drugs, expanded access programs, compassionate use programs, fast track regulations, accelerated approval, priority review, breakthrough drug designation, and use of "real-world evidence," and analogous initiatives elsewhere. On the other hand, efforts like the Right to Try Act may compromise the scientific rigor of the normal regulatory approach. The future is likely to see continued creative regulatory initiatives toward maintaining this balance.

There is also increased interest in encouraging the use of generic drugs, reducing costs, and reducing regulatory obstacles to the availability of generically equivalent drugs, once patents expire.

As much of drug development expands from just small molecules to include biologics, and as these mature on the market, regulators have had to develop a new framework to regulate biosimilars (see Chapter 23). The next few years are likely to see considerable new efforts in this vein.

Finally, there is an enormous increase in attention to drug safety, for example driven by drug safety issues identified with COX-2 inhibitors and even traditional nonsteroidal antiinflammatory drugs, and then by the thiazolidinediones, used for treatment of diabetes. The net result has been major regulatory change, and even new legislation. Between 2009 and 2012, for example, the FDA approved 110 new drugs and biologics for 120 indications, and only 13 of them did not have any postmarketing requirements [30].

The View from the Law

Finally, the importance of pharmacoepidemiology to the law has also been growing. The potential financial risk to drug manufacturers posed by lawsuits related to adverse drug effects is very large. Some financial payments have been enormous, and indeed put large multinational companies at risk. It is clear that the interest in the field and the need for more true experts in the field will increase accordingly.

Conclusion

> *There are no really "safe" biologically active drugs. There are only "safe" physicians.*
> (Harold A. Kaminetzsky, 1963)

All drugs have adverse effects. Pharmacoepidemiology will never succeed in preventing them. It can only detect them, hopefully early, and thereby educate healthcare providers and the public, which will lead to better medication use. Pharmacoepidemiology can also lead to safer use of medications through a better understanding of the factors that alter the risk/benefit balance of medications. The net results of increased activity in pharmacoepidemiology will be better for industry and academia but, most importantly, for the public's health. The next drug disaster cannot be prevented by pharmacoepidemiology but it can minimize its adverse public health impact by detecting it early. At the same time, it can improve the use of drugs that have a genuine role, protecting against the loss of useful drugs. The past few decades have demonstrated the utility of this new field. They also have pointed out some of its problems. With luck, the next few years will see the utility accentuated and the problems ameliorated.

References

1 Andrews EB, Avorn J, Bortnichak EA, *et al.* Guidelines for Good Epidemiology Practices for Drug, Device, and Vaccine Research in the United States. *Pharmacoepidemiol Drug Saf* 1996; **5**: 333–8.

2 Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2016; **25**: 2–10.

3 Strom BL, West SL, Sim E, Carson JL. The epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693–9.

4 Spielberg SP. Idiosyncratic drug reactions: interaction of development and genetics. *Semin Perinatol* 1992; **16**: 58–62.

5 Woosley RL, Drayer DE, Reidenberg MM, Nies AS, Carr K, Oates JA. Effect of

acetylator phenotype on the rate at which procainamide induces antinuclear antibodies and the lupus syndrome. *N Engl J Med* 1978; **298**: 1157–9.

6 Leeder JS, Riley RJ, Cook VA, Spielberg SP. Human anti-cytochrome P450 antibodies in aromatic anticonvulsant-induced hypersensitivity reactions. *J Pharmacol Exp Ther* 1992; **263**: 360–7.

7 Bray J, Sludden J, Griffin MJ, *et al*. Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide. *Br J Cancer* 2016; **102**: 1003–9.

8 Schmeling H, Biber D, Heins S, Horneff G. Influence of methylenetetrahydrofolate reductase polymorphisms on efficacy and toxicity of methotrexate in patients with juvenile idiopathic arthritis. *J Rheumatol* 2005; **32**: 1832–6.

9 Tassaneeyakul W, Jantararoungtong T, Chen P, *et al*. Strong association between HLA-B*5801 and allopurinol-induced Stevens–Johnson syndrome and toxic epidermal necrolysis in a Thai population. *Pharmacogenet Genom* 2009; **19**: 704–9.

10 Strom BL, Gibson GA. A systematic integrated approach to improving drug prescribing in an acute care hospital: a potential model for applied hospital pharmacoepidemiology. *Clin Pharmacol Ther* 1993; **54**: 126–33.

11 Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. *JAMA* 1991; **266**: 2847–51.

12 Bates DW, Leape LL, Cullen DJ, *et al*. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998; **280**: 1311–16.

13 Pullar T, Feely M. Problems of compliance with drug treatment: new solutions? *Pharm J* 1990; **245**: 213–15.

14 Urquhart J. The electronic medication event monitor – lessons for pharmacotherapy. *Clin Pharmacokinet* 1997; **32**: 345–56.

15 Joellenbeck LM, Zwanziger LL, Durch JS, Strom BL, eds. *The Anthrax Vaccine: Is It Safe? Does It Work?* Washington, DC: National Academies Press, 2002.

16 Committee on Smallpox Vaccination Program Implementation, Board on Health Promotion and Disease Prevention, Baciu A, Anason AP, Stratton K, Strom B (eds). *The Smallpox Vaccination Program: Public Health in an Age of Terrorism*. Washington, DC: National Academies Press, 2005.

17 Strom BL, Carson JL. Use of automated databases for pharmacoepidemiology research. *Epidemiol Rev* 1990; **12**: 87–107.

18 General Accountability Office. Testimony before the Special Committee on Aging, US Senate. Medicare Part D: Spending, beneficiary out-of-pocket costs, and efforts to obtain price concessions for certain high-cost drugs. GAO-10-529T. March 17, 2010. www.gao.gov/new.items/d10529t.pdf (accessed May 6, 2019).

19 Food and Drug Administration, eHealth Foundation, and the Brookings Institute. Sentinel Initiative: Structure, Function and Scope. www.fda.gov/media/76658/download (accessed May 6, 2019).

20 2008 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. http://www.cms.hhs.gov/ReportsTrustFunds/downloads/tr2008.pdf (accessed May 6, 2019).

21 Yudkin JS. The economics of pharmaceutical supply in Tanzania. *Int J Health Serv* 1980; **10**: 455–77.

22 Yudkin JS. Use and misuse of drugs in the Third World. *Dan Med Bull* 1984; **31(**Suppl 1): 11–17.

23 Lunde PKM. WHO's programme on essential drugs. Background, implementation, present state and prospectives. *Dan Med Bull* 1984; **31(**Suppl 1): 23–7.

24 Howard NJ, Laing RO. Changes in the World Health Organization essential drug list. *Lancet* 1991; **338**: 743–5.

25 Young FE. The role of the FDA in the effort against AIDS. *Public Health Rep* 1988; **103**: 242–5.

26 Joint Commission on Prescription Drug Use. *Report of the Joint Commission on Prescription Drug Use*. KF3869.R4. Rockville: Joint Commission on Prescription Drug Use, 1980.

27 Furberg CD, Levin AA, Gross PA, Shapiro RS, Strom BL. The FDA and drug safety: a proposal for sweeping changes. *Arch Intern Med* 2006; **166**: 1938–42.

28 Wood AJ, Stein CM, Woosley R. Making medicines safer – the need for an independent drug safety board. *N Engl J Med* 1998; **339**: 1851–4.

29 Moore TJ, Psaty BM, Furberg CD. Time to act on drug safety. *JAMA* 1998; **279**: 1571–3.

30 Wallach JD, Egilman AC, Dhruva SS, *et al*. Postmarket studies required by the US Food and Drug Administration for new drugs and biologics approved between 2009 and 2012: cross sectional analysis. *BMJ* 2018; **361**: k2031.

Appendix A

Sample Size Tables

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Table A1 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 970 717	2 788 497	6 306 290	29 429 320	37 837 603	10 510 431	3 153 120	1 634 946	1 051 034	756 742	583 904	394 133	211 445	142 27	61 134	22 318
0.00005	394 133	557 684	1 261 219	5 885 657	7 567 179	2 101 980	630 585	326 965	210 189	151 334	116 768	78 816	42 280	28 538	12 220	4458
0.0001	197 060	278 832	630 585	2 942 699	3 783 376	1 050 923	315 268	163 467	105 083	75 657	58 376	39 401	21 135	14 264	6106	2225
0.0005	39 401	55 751	126 078	588 332	756 333	210 078	63 015	32 669	20 999	15 117	11 662	7870	4219	2845	1215	439
0.001	19 694	27 865	63 015	294 037	377 953	104 973	31 483	16 320	10 488	7549	5823	3928	2104	1418	603	216
0.005	3928	5557	12 564	58 600	75 249	20 888	6257	3240	2080	1495	1152	775	412	276	114	37
0.01	1957	2769	6257	29 170	37 411	10 378	3104	1605	1028	738	568	381	201	133	53	15
0.05	381	538	1212	5627	7140	1969	582	297	188	133	101	65	32	19	4	—
0.10	184	259	582	2684	3357	918	266	133	82	57	42	26	10	4	—	—
0.15	118	166	372	1703	2095	568	161	79	47	32	23	13	—	—	—	—
0.20	85	120	266	1212	1465	393	109	52	30	19	13	6	—	—	—	—
0.25	65	92	203	918	1086	287	77	35	19	12	7	—	—	—	—	—
0.30	52	73	161	722	834	217	56	24	12	6	—	—	—	—	—	—
0.35	43	60	131	582	654	167	41	16	7	—	—	—	—	—	—	—
0.40	36	50	109	477	519	130	30	11	—	—	—	—	—	—	—	—
0.45	30	42	91	395	414	101	21	6	—	—	—	—	—	—	—	—
0.50	26	36	77	329	329	77	14	—	—	—	—	—	—	—	—	—
0.55	22	31	66	276	261	58	8	—	—	—	—	—	—	—	—	—
0.60	19	27	56	231	203	42	2	—	—	—	—	—	—	—	—	—
0.65	17	23	48	194	155	29	—	—	—	—	—	—	—	—	—	—
0.70	15	20	41	161	113	17	—	—	—	—	—	—	—	—	—	—
0.75	13	17	35	133	77	7	—	—	—	—	—	—	—	—	—	—
0.80	11	15	30	109	46	—	—	—	—	—	—	—	—	—	—	—
0.85	10	13	25	87	18	—	—	—	—	—	—	—	—	—	—	—
0.90	8	11	21	68	—	—	—	—	—	—	—	—	—	—	—	—
0.95	7	9	17	51	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : exposed ratio = 1 : 1. The sample size listed is the number of subjects needed in the exposed group. An equivalent number would be included in the control group.

Table A2 Sample size for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 529 057	2 153 636	4 825 616	22 279 822	28 149 090	7 764 537	2 302 889	1 183 563	755 529	540 883	415 381	278 329	147 626	99 000	41 938	15 197
0.0001	152 896	215 349	482 527	2 227 804	2 814 625	776 367	230 258	118 337	151 093	108 167	83 068	55 659	29 520	19 795	8384	3036
0.0005	30 570	43 057	96 475	445 402	562 673	155 196	46 024	23 651	75 539	54 077	41 528	27 825	14 756	9895	4189	1516
0.001	15 280	21 521	48 218	222 602	281 179	77 550	22 994	11 815	15 095	10 805	8297	5558	2946	1974	834	300
0.005	3047	4292	9613	44 362	55 984	15 433	4571	2346	7540	5396	4143	2774	1469	984	414	148
0.01	1518	2138	4787	22 082	27 834	7668	2268	1163	1496	1069	820	548	288	192	79	26
0.05	295	415	927	4258	5315	1456	426	216	740	528	404	269	141	93	37	11
0.10	142	200	444	2030	2500	680	196	97	136	95	72	47	23	14	3	—
0.15	91	128	283	1287	1561	421	119	58	60	41	31	19	8	3	—	—
0.20	66	92	203	916	1092	291	80	38	35	23	17	9	—	—	—	—
0.25	50	70	155	693	811	214	57	26	22	14	10	4	—	—	—	—
0.30	40	56	123	545	623	162	42	18	14	9	5	—	—	—	—	—
0.35	33	46	100	439	489	125	31	12	9	4	—	—	—	—	—	—
0.40	27	38	82	359	388	97	22	8	5	—	—	—	—	—	—	—
0.45	23	32	69	297	310	76	16	—	—	—	—	—	—	—	—	—
0.50	20	27	58	248	248	58	11	—	—	—	—	—	—	—	—	—
0.55	17	23	49	207	196	44	5	—	—	—	—	—	—	—	—	—
0.60	15	20	42	173	154	32	—	—	—	—	—	—	—	—	—	—
0.65	13	17	36	145	117	22	—	—	—	—	—	—	—	—	—	—
0.70	11	15	31	120	86	13	—	—	—	—	—	—	—	—	—	—
0.75	9	13	26	99	59	—	—	—	—	—	—	—	—	—	—	—
0.80	8	11	22	80	35	—	—	—	—	—	—	—	—	—	—	—
0.85	7	10	18	64	—	—	—	—	—	—	—	—	—	—	—	—
0.90	6	8	15	49	—	—	—	—	—	—	—	—	—	—	—	—
0.95	5	7	12	36	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : exposed ratio = 2 : 1. The sample size listed is the number of subjects needed in the exposed group. Double this number would be included in the control group.

Table A3 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1369471	1930847	4322614	19888657	24913372	6843626	2014756	1029014	653418	465696	356275	237254	124571	83030	34793	12510
0.00005	273886	386158	864495	3977589	4982452	1368657	402927	205788	130673	93131	71248	47445	24910	16602	6955	2499
0.0001	136938	193072	432230	1988706	2491087	684286	201449	102885	65330	46560	35619	23719	12452	8299	3476	1248
0.0005	27380	38603	86418	397599	497995	136790	40266	20563	13055	9303	7117	4738	2486	1656	692	247
0.001	13685	19294	43192	198711	248859	68352	20118	10272	6521	4646	3554	2365	1240	825	344	122
0.005	2729	3847	8611	39600	49549	13603	4000	2040	1294	921	703	467	244	161	66	21
0.01	1359	1916	4288	19711	24636	6759	1985	1011	640	455	347	230	119	78	31	9
0.05	264	372	830	3800	4705	1284	373	188	117	82	62	40	19	12	2	—
0.10	127	179	398	1811	2213	600	171	85	52	36	26	16	7	3	—	—
0.15	81	114	254	1148	1383	372	104	50	30	20	14	8	—	—	—	—
0.20	58	82	181	817	968	257	71	33	19	12	8	4	—	—	—	—
0.25	45	63	138	618	719	189	50	23	13	7	4	—	—	—	—	—
0.30	36	50	109	485	552	143	37	16	8	4	—	—	—	—	—	—
0.35	29	41	89	391	434	111	27	11	4	—	—	—	—	—	—	—
0.40	24	34	73	319	345	86	20	7	—	—	—	—	—	—	—	—
0.45	20	28	61	264	275	67	14	—	—	—	—	—	—	—	—	—
0.50	17	24	52	220	220	52	9	—	—	—	—	—	—	—	—	—
0.55	15	21	44	184	175	39	—	—	—	—	—	—	—	—	—	—
0.60	13	18	37	154	137	29	—	—	—	—	—	—	—	—	—	—
0.65	11	15	32	128	105	19	—	—	—	—	—	—	—	—	—	—
0.70	10	13	27	106	77	10	—	—	—	—	—	—	—	—	—	—
0.75	8	11	23	87	53	—	—	—	—	—	—	—	—	—	—	—
0.80	7	10	19	71	31	—	—	—	—	—	—	—	—	—	—	—
0.85	6	8	16	56	—	—	—	—	—	—	—	—	—	—	—	—
0.90	5	7	13	43	—	—	—	—	—	—	—	—	—	—	—	—
0.95	4	6	11	31	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : exposed ratio = 3 : 1. The sample size listed is the number of subjects needed in the exposed group. Triple this number would be included in the control group.

Table A4 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1285566	1815876	4068209	18690665	23293643	6381472	1869238	950463	601217	427061	325766	215895	112429	74554	30945	11048
0.00005	257106	363164	813616	3737999	4658521	1276231	373825	190079	120234	85404	65147	43174	22482	14907	6186	2207
0.0001	128548	181575	406791	1868916	2329131	638076	186899	95031	60111	42697	32569	21583	11238	7451	3091	1102
0.0005	25702	36304	81332	373649	465619	127552	37358	18993	12013	8532	6507	4311	2244	1487	615	218
0.001	12846	18145	40650	186741	232680	63737	18665	9488	6000	4261	3249	2152	1119	741	306	107
0.005	2562	3618	8104	37214	46329	12684	3711	1884	1190	844	643	425	220	145	58	19
0.01	1276	1802	4035	18523	23035	6303	1842	934	589	417	318	209	107	70	27	8
0.05	248	349	781	3571	4399	1198	346	174	108	76	57	36	17	10	2	—
0.10	119	168	374	1702	2070	560	159	78	48	33	24	15	6	2	—	—
0.15	76	107	238	1079	1294	347	97	47	28	19	13	7	—	—	—	—
0.20	55	77	171	767	905	240	66	31	18	11	8	3	—	—	—	—
0.25	42	59	130	580	672	177	47	21	12	7	4	—	—	—	—	—
0.30	33	47	103	456	517	134	34	15	7	—	—	—	—	—	—	—
0.35	27	38	83	366	406	103	25	10	3	—	—	—	—	—	—	—
0.40	23	32	69	300	323	81	18	6	—	—	—	—	—	—	—	—
0.45	19	27	58	248	258	63	13	—	—	—	—	—	—	—	—	—
0.50	16	23	48	206	206	48	8	—	—	—	—	—	—	—	—	—
0.55	14	19	41	172	164	37	—	—	—	—	—	—	—	—	—	—
0.60	12	16	35	144	128	27	—	—	—	—	—	—	—	—	—	—
0.65	10	14	30	120	98	18	—	—	—	—	—	—	—	—	—	—
0.70	9	12	25	99	72	7	—	—	—	—	—	—	—	—	—	—
0.75	8	10	21	81	50	—	—	—	—	—	—	—	—	—	—	—
0.80	6	9	18	66	29	—	—	—	—	—	—	—	—	—	—	—
0.85	6	8	15	52	—	—	—	—	—	—	—	—	—	—	—	—
0.90	5	6	12	39	—	—	—	—	—	—	—	—	—	—	—	—
0.95	4	5	10	28	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : exposed ratio = 4 : 1. The sample size listed is the number of subjects needed in the exposed group. Quadruple this number would be included in the control group.

Table A5 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 472 091	2 082 958	4 710 686	21 983 178	28 264 016	7 851 105	2 355 325	1 221 276	785 104	565 273	436 166	294 411	157 946	106 615	45 666	16 672
0.00005	294 411	416 580	942 108	4 396 481	5 652 548	1 570 142	471 036	244 238	157 008	113 044	87 224	58 875	31 583	21 318	9129	3330
0.0001	147 201	208 283	471 036	2 198 144	2 826 115	785 022	235 500	122 108	78 496	56 515	43 606	29 433	15 788	10 656	4562	1663
0.0005	29 433	41 645	94 178	439 474	564 968	156 925	47 071	24 404	15 686	11 292	8712	5879	3152	2126	908	329
0.001	14 711	20 816	47 071	219 641	282 325	78 413	23 518	12 191	7835	5639	4350	2935	1572	1060	451	162
0.005	2935	4152	9385	43 774	56 210	15 604	4675	2421	1554	1117	861	579	309	207	86	28
0.01	1463	2069	4675	21 790	27 946	7752	2319	1199	769	552	425	285	151	100	40	12
0.05	285	402	906	4204	5334	1471	435	222	141	100	76	49	24	15	3	—
0.10	138	194	435	2005	2508	686	200	100	62	43	32	20	8	4	—	—
0.15	89	125	278	1273	1566	425	121	59	36	24	17	10	—	—	—	—
0.20	64	90	200	906	1095	294	82	39	15	15	10	5	—	—	—	—
0.25	49	69	152	686	812	215	58	27	10	9	6	—	—	—	—	—
0.30	40	55	121	540	623	163	42	19	6	5	—	—	—	—	—	—
0.35	33	45	99	435	489	125	31	13	—	—	—	—	—	—	—	—
0.40	27	38	82	357	388	97	23	8	—	—	—	—	—	—	—	—
0.45	23	32	69	295	309	76	16	5	—	—	—	—	—	—	—	—
0.50	20	27	58	247	247	58	11	—	—	—	—	—	—	—	—	—
0.55	17	24	50	207	195	44	7	—	—	—	—	—	—	—	—	—
0.60	15	20	42	173	152	32	2	—	—	—	—	—	—	—	—	—
0.65	13	18	36	145	116	22	—	—	—	—	—	—	—	—	—	—
0.70	11	15	31	121	85	13	—	—	—	—	—	—	—	—	—	—
0.75	10	13	27	100	58	6	—	—	—	—	—	—	—	—	—	—
0.80	9	12	23	82	35	—	—	—	—	—	—	—	—	—	—	—
0.85	8	10	19	66	14	—	—	—	—	—	—	—	—	—	—	—
0.90	7	9	16	51	—	—	—	—	—	—	—	—	—	—	—	—
0.95	6	8	14	38	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : exposed ratio = 1 : 1. The sample size listed is the number of subjects needed in the exposed group. An equivalent number would be included in the control group.

Table A6 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 190 356	1 663 432	3 680 447	16 792 779	20 878 641	5 726 194	1 683 582	859 799	546 209	389 547	298 242	198 909	104 767	69 986	29 458	10 630
0.00005	238 065	332 677	736 066	3 358 436	4 175 543	1 145 183	336 697	171 948	109 233	77 903	59 643	39 777	20 950	13 994	5889	2124
0.0001	119 028	166 332	368 018	1 679 143	2 087 655	572 556	168 336	85 967	54 611	38 947	29 818	19 886	10 473	6995	2943	1061
0.0005	23 799	33 257	73 580	335 708	417 346	114 455	33 648	17 182	10 914	7783	5958	3973	2091	1396	586	210
0.001	11 895	16 622	36 775	167 779	208 557	57 193	16 812	8584	5452	3887	2975	1983	1043	696	292	104
0.005	2372	3315	7332	33 436	41 526	11 382	3343	1705	1082	771	589	392	205	136	56	19
0.01	1182	1651	3651	16 643	20 647	5656	1659	845	536	381	291	193	100	66	26	8
0.05	230	321	707	3208	3944	1075	312	157	99	69	52	34	17	10	2	—
0.10	111	154	339	1529	1856	503	144	71	44	30	23	14	6	3	—	—
0.15	71	99	216	969	1160	312	88	43	26	17	13	7	—	—	—	—
0.20	51	71	155	689	812	216	60	28	17	11	8	4	—	—	—	—
0.25	39	54	118	522	603	159	43	20	11	7	4	—	—	—	—	—
0.30	31	43	93	410	464	121	32	14	7	4	—	—	—	—	—	—
0.35	26	35	76	330	365	93	23	10	4	—	—	—	—	—	—	—
0.40	21	29	63	270	290	73	17	6	—	—	—	—	—	—	—	—
0.45	18	25	52	223	232	57	13	—	—	—	—	—	—	—	—	—
0.50	15	21	44	186	186	44	9	—	—	—	—	—	—	—	—	—
0.55	13	18	38	155	148	34	5	—	—	—	—	—	—	—	—	—
0.60	11	16	32	130	116	25	—	—	—	—	—	—	—	—	—	—
0.65	10	13	27	108	89	18	—	—	—	—	—	—	—	—	—	—
0.70	9	12	23	90	66	11	—	—	—	—	—	—	—	—	—	—
0.75	8	10	20	74	46	—	—	—	—	—	—	—	—	—	—	—
0.80	7	9	17	60	28	—	—	—	—	—	—	—	—	—	—	—
0.85	6	7	14	47	—	—	—	—	—	—	—	—	—	—	—	—
0.90	5	6	12	36	—	—	—	—	—	—	—	—	—	—	—	—
0.95	4	5	9	26	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : exposed ratio = 2 : 1. The sample size listed is the number of subjects needed in the exposed group. Double this number would be included in the control group.

Table A7 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 088 323	1 516 254	3 330 831	15 057 392	18 412 768	5 014 203	1 456 566	736 622	464 207	328 848	250 342	165 451	85 870	56 861	23 565	8410
0.00005	217 658	303 242	666 145	3 011 370	3 682 391	1 002 792	291 297	147 315	92 835	65 764	50 064	33 087	17 171	11 370	4711	1681
0.0001	108 825	151 615	333 059	1 505 617	1 841 094	501 366	145 638	73 651	46 413	32 879	25 029	16 541	8584	5684	2355	839
0.0005	21 759	30 314	66 590	301 015	368 057	100 225	29 111	14 721	9276	6570	5001	3305	1714	1134	469	166
0.001	10 875	15 151	33 281	150 439	183 927	50 082	14 545	7354	4634	3282	2498	1650	855	566	233	82
0.005	2169	3021	6635	29 979	36 623	9968	2892	1461	920	651	495	326	168	111	45	15
0.01	1080	1505	3304	14 922	18 210	4954	1436	725	456	322	245	161	83	54	21	6
0.05	210	292	639	2876	3480	942	271	135	84	59	44	29	14	8	2	—
0.10	101	140	306	1370	1638	441	125	62	38	26	19	12	5	2	—	—
0.15	65	90	195	868	1025	274	76	37	22	15	11	6	—	—	—	—
0.20	46	64	139	617	718	190	52	25	14	9	6	3	—	—	—	—
0.25	36	49	106	466	534	140	37	17	10	6	4	—	—	—	—	—
0.30	28	39	84	366	411	107	28	12	6	3	—	—	—	—	—	—
0.35	23	32	68	294	323	83	21	9	4	—	—	—	—	—	—	—
0.40	19	26	56	240	257	65	15	6	—	—	—	—	—	—	—	—
0.45	16	22	47	199	206	51	11	—	—	—	—	—	—	—	—	—
0.50	14	19	39	165	165	39	8	—	—	—	—	—	—	—	—	—
0.55	12	16	33	138	132	30	—	—	—	—	—	—	—	—	—	—
0.60	10	14	28	115	104	23	—	—	—	—	—	—	—	—	—	—
0.65	9	12	24	96	80	16	—	—	—	—	—	—	—	—	—	—
0.70	8	10	20	79	60	9	—	—	—	—	—	—	—	—	—	—
0.75	7	9	17	65	42	—	—	—	—	—	—	—	—	—	—	—
0.80	6	7	14	52	26	—	—	—	—	—	—	—	—	—	—	—
0.85	5	6	12	41	—	—	—	—	—	—	—	—	—	—	—	—
0.90	4	5	10	31	—	—	—	—	—	—	—	—	—	—	—	—
0.95	3	4	8	22	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : exposed ratio = 3 : 1. The sample size listed is the number of subjects needed in the exposed group. Triple this number would be included in the control group.

Table A8 Sample sizes for cohort studies.

Incidence in control group	Relative risk to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 034 606	1 440 316	3 154 116	14 188 116	17 178 604	4 657 092	1 342 104	674 194	422 454	297 814	225 764	148 182	76 019	49 975	20 438	7223
0.00005	206 915	288 054	630 802	2 837 520	3 435 570	931 374	268 406	134 830	84 485	59 558	45 149	29 633	15 201	9993	4086	1443
0.0001	103 454	144 022	315 388	1 418 696	1 717 691	465 659	134 194	67 410	42 238	29 776	22 572	14 815	7599	4995	2042	721
0.0005	20 685	28 795	63 057	283 636	343 387	93 087	26 824	13 473	8442	5950	4510	2960	1518	997	407	143
0.001	10 338	14 392	31 515	141 754	171 599	46 516	13 402	6731	4217	2972	2253	1478	757	497	203	71
0.005	2061	2870	6282	28 248	34 169	9259	2665	1338	837	590	446	292	149	98	39	13
0.01	027	1429	3128	14 059	16 990	4601	1323	663	415	292	221	144	73	48	19	6
0.05	199	277	605	2709	3247	876	250	124	77	53	40	26	12	8	2	—
0.10	96	133	289	1290	1529	410	115	57	35	24	17	11	5	2	—	—
0.15	61	85	184	817	957	255	71	34	20	14	10	6	—	—	—	—
0.20	44	61	132	581	670	177	48	23	13	9	6	3	—	—	—	—
0.25	34	47	100	439	499	130	35	16	9	5	3	—	—	—	—	—
0.30	27	37	79	344	384	99	26	11	6	—	—	—	—	—	—	—
0.35	22	30	64	277	302	77	19	8	3	—	—	—	—	—	—	—
0.40	18	25	53	226	241	60	14	5	—	—	—	—	—	—	—	—
0.45	15	21	44	186	193	47	10	—	—	—	—	—	—	—	—	—
0.50	13	18	37	155	155	37	7	—	—	—	—	—	—	—	—	—
0.55	11	15	31	129	124	28	—	—	—	—	—	—	—	—	—	—
0.60	9	13	26	108	97	21	—	—	—	—	—	—	—	—	—	—
0.65	8	11	22	89	75	15	—	—	—	—	—	—	—	—	—	—
0.70	7	9	19	74	56	7	—	—	—	—	—	—	—	—	—	—
0.75	6	8	16	60	39	—	—	—	—	—	—	—	—	—	—	—
0.80	5	7	13	48	24	—	—	—	—	—	—	—	—	—	—	—
0.85	4	6	11	38	—	—	—	—	—	—	—	—	—	—	—	—
0.90	4	5	9	28	—	—	—	—	—	—	—	—	—	—	—	—
0.95	3	4	7	20	—	—	—	—	—	—	—	—	—	—	—	—

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : exposed ratio = 4 : 1. The sample size listed is the number of subjects needed in the exposed group. Quadruple this number would be included in the control group.

Table A9 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 970 728	2 788 519	6 306 363	29 429 793	37 838 497	10 510 715	3 153 225	1 635 011	1 051 081	756 780	583 937	394 159	211 464	142 743	61 147	22 330
0.00005	394 143	557 705	1 261 292	5 886 130	7 568 072	2 102 264	630 690	327 029	210 236	151 372	116 801	78 842	42 300	28 555	12 234	4469
0.0001	197 070	278 853	630 659	2 943 172	3 784 269	1 051 207	315 373	163 532	105 130	75 696	58 409	39 427	21 155	14 281	6120	2237
0.0005	39 412	55 772	126 151	588 806	757 227	210 362	63 120	32 734	21 046	15 155	11 695	7896	4238	2862	1228	451
0.001	19 704	27 887	63 088	294 510	378 847	105 257	31 588	16 384	10 535	7587	5856	3954	2124	1435	617	228
0.005	3939	5579	12 638	59 074	76 145	21 173	6363	3304	2127	1533	1184	801	432	293	128	49
0.01	1968	2790	6331	29 646	38 309	10 663	3210	1669	1076	777	601	407	221	150	67	27
0.05	391	560	1288	6111	8059	2261	690	363	237	172	135	93	52	37	18	9
0.10	195	281	659	3181	4302	1219	379	202	133	98	77	54	32	23	13	8
0.15	129	189	451	2215	3072	879	278	150	100	75	60	43	26	19	11	8
0.20	97	143	348	1741	2476	716	230	126	85	64	52	37	23	18	11	8
0.25	77	116	287	1465	2137	624	203	113	77	59	48	35	23	18	12	9
0.30	64	98	248	1289	1930	569	188	106	73	56	46	34	23	18	13	10
0.35	56	86	222	1174	1802	536	180	103	72	56	46	35	24	19	14	11
0.40	49	77	203	1097	1727	519	177	102	72	56	47	36	25	20	15	12
0.45	44	70	191	1048	1694	513	178	104	74	58	49	38	27	22	17	14
0.50	40	66	182	1023	1696	519	182	108	77	61	52	40	29	24	19	16
0.55	38	62	178	1019	1732	535	191	114	82	66	56	44	32	27	21	18
0.60	36	61	177	1035	1806	562	203	123	89	72	61	49	36	31	25	21
0.65	35	60	180	1077	1927	605	222	135	99	80	69	56	42	36	29	25
0.70	34	61	188	1149	2110	669	248	153	113	92	79	64	49	43	35	31
0.75	35	64	203	1268	2390	764	287	178	133	109	94	77	59	52	43	38
0.80	37	70	230	1465	2831	913	348	218	164	135	117	97	75	66	55	49
0.85	43	82	278	1811	3591	1168	451	285	216	179	156	129	101	90	75	68
0.90	54	108	379	2527	5143	1687	659	420	320	266	233	195	154	137	116	105
0.95	93	190	690	4717	9851	3257	1288	828	635	531	466	391	313	280	238	217

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : case ratio = 1 : 1. The sample size listed is the number of subjects needed in the case group. An equivalent number would be included in the control group.

Table A10 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 529 065	2 153 652	4 825 672	22 280 178	28 149 758	7 764 749	2 302 966	1 183 610	755 564	540 911	415 405	278 348	147 639	99 012	41 948	15 205
0.00005	305 811	430 731	965 148	4 456 162	5 630 233	1 553 041	460 628	236 743	151 128	108 194	83 091	55 678	29 534	19 807	8393	3044
0.0001	152 904	215 366	482 583	2 228 160	2 815 293	776 578	230 335	118 385	75 573	54 105	41 552	27 844	14 770	9906	4199	1524
0.0005	30 578	43 073	96 531	445 759	563 340	155 407	46 101	23 698	15 130	10 833	8321	5577	2960	1986	843	307
0.001	15 288	21 537	48 274	222 959	281 846	77 761	23 072	11 862	7574	5424	4167	2793	1483	996	424	155
0.005	3055	4308	9669	44 719	56 653	15 644	4649	2393	1530	1097	844	567	302	204	88	34
0.01	1526	2154	4843	22 440	28 505	7880	2346	1210	775	556	428	289	155	105	46	19
0.05	303	431	984	4623	6001	1674	506	264	171	124	97	66	37	26	13	7
0.10	150	216	503	2405	3207	904	279	148	97	71	56	39	23	17	9	6
0.15	100	145	343	1673	2292	653	205	111	74	55	44	31	19	14	8	6
0.20	74	110	265	1313	1849	533	170	93	63	47	38	28	17	13	8	6
0.25	59	89	218	1104	1597	465	151	84	57	44	35	26	17	13	9	6
0.30	49	75	188	971	1443	425	140	79	55	42	34	26	17	14	9	7
0.35	42	65	168	883	1349	401	135	77	54	42	34	26	18	14	10	8
0.40	37	58	154	825	1294	388	133	77	54	42	35	27	19	15	11	9
0.45	33	53	144	788	1270	385	133	78	56	44	37	28	20	17	13	10
0.50	31	50	137	768	1272	389	137	81	58	46	39	31	22	19	14	12
0.55	28	47	133	764	1301	402	144	86	62	50	42	33	24	21	16	14
0.60	27	45	133	775	1357	423	154	93	68	55	47	37	28	24	19	16
0.65	26	45	135	805	1449	456	168	103	76	61	52	42	32	28	22	19
0.70	26	45	140	859	1588	505	188	116	86	70	61	49	38	33	27	23
0.75	26	47	151	947	1799	577	218	136	102	84	72	59	46	40	33	29
0.80	28	51	170	1092	2133	690	265	166	125	104	90	74	58	51	42	38
0.85	31	60	205	1349	2708	884	343	218	165	137	120	100	78	70	58	53
0.90	39	78	279	1880	3881	1278	503	322	246	205	180	150	119	107	90	82
0.95	66	137	506	3505	7438	2472	984	635	489	410	360	303	243	218	186	169

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : case ratio = 2 : 1. The sample size listed is the number of subjects needed in the case group. Double this number would be included in the control group.

Table A11 Sample size for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 369 478	1 930 861	4 322 663	19 888 975	24 913 964	6 843 813	2 014 824	1 029 056	653 448	465 720	356 295	237 271	124 583	83 040	34 800	12 517
0.00005	273 893	386 172	864 545	3 977 907	4 983 044	1 368 844	402 996	205 830	130 703	93 155	71 268	47 461	24 922	16 612	6963	2506
0.0001	136 945	193 086	432 280	1 989 023	2 491 679	684 473	201 517	102 927	65 360	46 584	35 640	23 735	12 464	8309	3483	1254
0.0005	27 387	38 617	86 468	397 917	498 587	136 977	40 334	20 604	13 086	9328	7137	4754	2498	1666	700	253
0.001	13 692	19 309	43 242	199 028	249 451	68 540	20 186	10 314	6551	4671	3574	2382	1252	836	352	128
0.005	2736	3862	8661	39 918	50 143	13 790	4068	2082	1324	945	724	484	256	171	73	28
0.01	1367	1931	4338	20 030	25 231	6947	2054	1053	671	480	368	246	131	88	39	16
0.05	271	387	881	4125	5313	1477	444	231	149	108	84	57	32	22	11	6
0.10	134	194	450	2145	2841	799	245	129	85	62	49	34	20	14	8	5
0.15	89	130	307	1491	2031	577	180	97	64	48	38	27	16	12	7	5
0.20	66	98	236	1171	1639	471	150	82	55	41	33	24	15	12	7	5
0.25	53	79	195	984	1417	412	133	74	50	38	31	23	15	12	8	6
0.30	44	67	168	865	1281	376	124	70	48	37	30	23	15	12	8	6
0.35	38	58	150	786	1197	355	119	68	47	37	30	23	16	13	9	7
0.40	33	52	137	734	1149	345	118	68	48	37	31	24	16	14	10	8
0.45	30	47	128	700	1128	342	119	69	49	39	32	25	18	15	11	9
0.50	27	44	122	682	1131	346	122	72	52	41	35	27	19	16	12	10
0.55	25	42	119	679	1156	357	128	76	55	44	38	30	22	18	14	12
0.60	24	40	118	689	1207	377	137	83	60	49	41	33	25	21	17	14
0.65	23	40	119	715	1289	406	150	91	67	55	47	38	28	24	20	17
0.70	23	40	124	762	1414	450	168	104	77	63	54	44	33	29	24	21
0.75	23	42	133	839	1602	515	195	121	91	75	65	53	41	36	29	26
0.80	24	45	150	968	1900	616	236	149	112	93	80	66	52	45	38	34
0.85	27	52	180	1194	2413	789	307	195	148	123	107	89	70	62	52	46
0.90	34	68	245	1664	3459	1142	450	288	220	184	161	134	107	95	80	72
0.95	57	119	444	3100	6632	2208	881	569	438	367	323	271	217	194	165	150

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control:case ratio = 3:1. The sample size listed is the number of subjects needed in the case group. Triple this number would be included in the control group.

Table A12 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 285 573	1 815 890	4 068 256	18 690 963	23 294 197	6 381 647	1 869 301	950 501	601 245	427 084	325 786	215 910	112 440	74 563	30 952	11 054
0.00005	257 112	363 178	813 662	3 738 297	4 659 075	1 276 406	373 889	190 118	120 262	85 427	65 166	43 189	22 493	14 916	6193	2213
0.0001	128 555	181 589	406 838	1 869 214	2 329 685	638 251	186 963	95 070	60 139	42 720	32 588	21 599	11 249	7461	3098	1108
0.0005	25 709	36 318	81 379	373 947	466 173	127 727	37 422	19 032	12 041	8554	6526	4326	2255	1496	622	224
0.001	12 853	18 159	40 697	187 039	233 234	63 912	18 729	9527	6028	4284	3269	2167	1130	750	313	113
0.005	2568	3632	8151	37 513	46 884	12 860	3775	1923	1219	867	662	440	231	154	65	25
0.01	1283	1816	4082	18 823	23 592	6479	1906	973	618	440	337	224	118	79	34	14
0.05	255	363	829	3876	4969	1378	412	214	137	99	77	52	29	20	10	5
0.10	126	182	423	2015	2658	746	228	120	78	57	45	31	18	13	7	4
0.15	83	122	289	1401	1901	539	168	90	60	44	35	25	15	11	7	4
0.20	62	92	222	1099	1534	440	140	76	51	38	31	22	14	11	7	5
0.25	50	74	183	923	1326	385	125	69	47	36	29	21	14	11	7	5
0.30	41	63	158	812	1200	352	116	65	45	34	28	21	14	11	7	6
0.35	35	55	140	738	1122	333	111	63	44	34	28	21	14	12	8	6
0.40	31	49	128	688	1077	323	110	63	45	35	29	22	15	13	9	7
0.45	28	44	120	657	1058	320	111	65	46	36	30	23	17	14	10	8
0.50	25	41	114	640	1060	324	114	67	48	38	32	25	18	15	11	10
0.55	23	39	111	636	1084	335	120	72	52	41	35	28	20	17	13	11
0.60	22	38	110	645	1132	354	128	78	57	46	39	31	23	20	15	13
0.65	21	37	111	669	1209	381	140	86	63	51	44	35	26	23	18	16
0.70	21	37	116	713	1326	422	158	97	72	59	51	41	31	27	22	19
0.75	21	39	125	786	1504	483	183	114	85	70	61	50	38	33	27	24
0.80	22	42	140	905	1784	579	222	140	105	87	75	62	48	42	35	31
0.85	25	48	168	1117	2266	742	289	183	139	115	101	83	65	58	48	43
0.90	31	63	228	1556	3248	1073	423	271	207	173	151	126	100	89	75	67
0.95	52	110	412	2897	6229	2076	829	536	412	345	303	255	203	182	154	139

$\alpha = 0.05$ (two-tailed); $\beta = 0.10$ (power = 90%); control : case ratio = 4 : 1. The sample size listed is the number of subjects needed in the case group. Quadruple this number would be included in the control group.

Table A13 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 472 099	2 082 974	4 710 741	21 983 531	28 264 683	7 851 317	2 355 404	1 221 324	785 139	565 302	436 191	294 430	157 960	106 627	45 676	16 681
0.00005	294 418	416 596	942 163	4 396 835	5 653 216	1 570 354	471 115	244 286	157 043	113 073	87 248	58 894	31 598	21 330	9139	3339
0.0001	147 208	208 299	471 091	2 198 497	2 826 782	785 234	235 579	122 156	78 531	56 544	43 631	29 452	15 803	10 668	4572	1671
0.0005	29 440	41 661	94 233	439 828	565 636	157 137	47 150	24 452	15 721	11 321	8736	5899	3166	2138	918	337
0.001	14 719	20 831	47 126	219 994	282 992	78 625	23 596	12 239	7870	5668	4375	2954	1587	1072	461	171
0.005	2943	4168	9441	44 128	56 879	15 816	4753	2469	1589	1146	885	599	323	219	96	37
0.01	1470	2085	4730	22 145	28 617	7966	2398	1248	804	581	449	305	165	113	50	20
0.05	293	419	962	4566	6020	1690	516	272	177	129	101	70	39	28	14	7
0.10	146	211	493	2377	3214	911	283	151	100	74	58	41	24	18	10	6
0.15	97	142	337	1655	2295	657	208	113	75	56	45	32	20	15	9	6
0.20	73	107	260	1301	1850	535	172	95	64	48	39	28	18	14	9	6
0.25	58	87	215	1095	1597	466	152	85	58	44	36	27	17	14	9	7
0.30	49	74	186	964	1442	425	141	80	55	42	35	26	18	14	10	8
0.35	42	65	166	877	1346	401	135	77	54	42	35	26	18	15	11	9
0.40	37	58	152	820	1291	388	133	77	54	42	35	27	19	16	12	10
0.45	33	53	143	784	1266	384	133	78	56	44	37	29	20	17	13	11
0.50	31	50	137	765	1267	388	137	81	58	46	39	31	22	19	15	12
0.55	29	47	133	761	1294	400	143	85	62	50	42	33	25	21	16	14
0.60	27	46	133	774	1350	421	152	92	67	54	46	37	28	24	19	16
0.65	26	45	135	805	1440	453	166	101	75	61	52	42	32	27	22	19
0.70	26	46	141	859	1577	500	186	115	85	69	60	49	37	32	26	23
0.75	27	48	152	948	1785	571	215	134	100	82	71	58	45	39	33	29
0.80	28	53	172	1095	2115	682	260	163	123	101	88	73	57	50	42	37
0.85	32	62	208	1353	2683	873	337	213	162	134	117	97	76	68	57	51
0.90	41	81	283	1888	3842	1260	493	314	240	200	175	146	116	103	87	79
0.95	70	142	516	3524	7359	2433	962	619	475	397	349	293	234	210	179	162

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : case ratio = 1 : 1. The sample size listed is the number of subjects needed in the case group. An equivalent number would be included in the control group.

Table A14 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 190 363	1 663 444	3 680 489	16 793 046	20 879 138	5 726 351	1 683 639	859 834	546 235	389 568	298 260	198 923	104 777	69 995	29 465	10 635
0.00005	238 071	332 689	736 108	3 358 703	4 176 039	1 145 339	336 754	171 983	109 259	77 923	59 660	39 791	20 960	14 003	5896	2129
0.0001	119 034	166 344	368 060	1 679 410	2 088 152	572 713	168 393	86 001	54 637	38 967	29 835	19 899	10 483	7004	2950	1066
0.0005	23 805	33 269	73 622	335 976	417 842	114 612	33 705	17 216	10 939	7803	5975	3986	2101	1405	593	216
0.001	11 901	16 635	36 817	168 047	209 054	57 349	16 869	8618	5477	3907	2993	1997	1053	705	298	109
0.005	2378	3327	7374	33 704	42 024	11 540	3400	1740	1107	791	607	406	215	145	63	24
0.01	1188	1664	3693	16 911	21 146	5814	1717	880	561	402	308	207	211	75	33	14
0.05	236	333	750	3482	4455	1237	371	193	125	91	70	48	27	19	10	5
0.10	117	167	383	1810	2383	669	205	109	71	53	41	29	17	13	7	5
0.15	77	112	261	1258	1704	484	152	82	54	41	32	23	14	11	7	5
0.20	58	84	201	987	1376	396	126	69	47	35	28	21	13	10	7	5
0.25	46	68	166	829	1190	346	112	62	43	33	27	20	13	10	7	5
0.30	38	57	143	729	1076	316	105	59	41	32	26	20	13	11	7	6
0.35	33	50	127	662	1006	299	101	58	40	31	26	20	14	11	8	7
0.40	29	45	116	618	966	290	99	58	41	32	27	21	15	12	9	7
0.45	26	41	108	590	949	288	100	59	42	33	28	22	16	13	10	8
0.50	24	38	103	574	951	292	103	61	44	35	30	24	17	15	11	10
0.55	22	36	100	571	973	301	108	65	47	38	32	26	19	16	13	11
0.60	21	34	99	579	1016	318	116	70	52	42	36	29	22	19	15	13
0.65	20	34	101	601	1085	343	127	78	58	47	40	33	25	22	18	16
0.70	20	34	105	640	1190	380	143	89	66	54	47	38	29	26	21	19
0.75	20	35	112	705	1350	435	166	104	78	64	56	46	36	32	26	23
0.80	21	38	126	812	1601	520	201	127	96	80	70	58	45	40	34	30
0.85	23	44	152	1002	2034	667	261	167	127	106	93	77	61	55	46	42
0.90	29	58	205	1395	2916	965	383	246	189	158	139	117	94	84	71	65
0.95	48	100	371	2598	5592	1868	750	487	376	316	279	236	190	171	147	134

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : case ratio = 2 : 1. The sample size listed is the number of subjects needed in the case group. Double this number would be included in the control group.

Table A15 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 088 329	1 516 265	3 330 869	15 057 631	18 413 208	5 014 341	1 456 616	736 652	464 229	328 865	250 357	165 463	85 879	56 868	23 570	8415
0.00005	217 664	303 253	666 182	3 011 608	3 682 831	1 002 930	291 347	147 345	92 856	65 782	50 079	33 098	17 180	11 377	4717	1685
0.0001	108 831	151 626	333 096	1 505 856	1 841 534	501 504	145 688	73 681	46 435	32 896	25 044	16 553	8592	5691	2360	844
0.0005	21 764	30 325	66 628	301 253	368 496	100 363	29 161	14 751	9298	6588	5016	3316	1723	1141	474	171
0.001	10 881	15 162	33 319	150 678	184 367	50 220	14 595	7384	4655	3299	2513	1662	864	573	239	87
0.005	2174	3032	6672	30 218	37 064	10 107	2943	1491	942	668	510	338	177	118	50	19
0.01	1086	1516	3342	15 161	18 652	5093	1486	755	478	340	259	173	91	61	27	11
0.05	215	303	678	3120	3932	1085	323	167	107	77	60	41	23	16	8	4
0.10	107	152	345	1620	2105	588	179	94	62	45	35	25	15	11	6	4
0.15	70	101	235	1125	1507	426	132	71	47	35	28	20	12	9	6	4
0.20	52	76	181	882	1218	349	111	60	41	31	25	18	11	9	6	4
0.25	42	62	149	741	1053	305	99	55	37	29	23	17	11	9	6	5
0.30	35	52	128	650	954	280	92	52	36	28	23	17	12	9	7	5
0.35	30	45	114	590	892	265	89	51	36	28	23	18	12	10	7	6
0.40	26	40	104	550	857	257	88	51	36	28	24	18	13	11	8	7
0.45	23	36	97	525	843	256	89	52	37	30	25	19	14	12	9	7
0.50	21	34	92	511	846	259	92	55	39	32	27	21	15	13	10	9
0.55	19	32	89	507	866	268	97	58	42	34	29	23	17	15	12	10
0.60	18	30	88	514	905	283	104	63	46	38	32	26	19	17	13	12
0.65	18	30	89	533	967	306	114	70	52	42	36	30	23	20	16	14
0.70	17	30	92	567	1061	339	128	80	60	49	42	35	27	23	19	17
0.75	17	31	99	624	1204	389	149	93	70	58	51	42	32	29	24	21
0.80	18	33	111	718	1429	466	181	115	87	72	63	52	41	37	31	28
0.85	20	38	132	884	1817	598	235	151	115	96	84	70	56	50	42	38
0.90	25	50	179	1230	2607	867	345	223	172	144	127	107	85	77	65	59
0.95	41	85	323	2288	5002	1678	678	442	342	288	255	215	174	157	134	123

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : case ratio = 3 : 1. The sample size listed is the number of subjects needed in the case group. Triple this number would be included in the control group.

Table A16 Sample sizes for case–control studies.

Prevalence in control group	Odds ratio to be detected															
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0
0.00001	1 034 611	1 440 327	3 154 151	14 188 340	17 179 015	4 657 221	1 342 151	674 222	422 474	297 830	225 778	148 193	76 026	49 982	20 443	7227
0.00005	206 920	288 065	630 838	2 837 745	3 435 981	931 503	268 452	134 858	84 505	59 574	45 162	29 644	15 209	9999	4091	1447
0.0001	103 459	144 032	315 424	1 418 920	1 718 102	465 788	134 240	67 438	42 259	29 792	22 585	14 825	7607	5002	2047	725
0.0005	20 690	28 806	63 092	283 861	343 799	93 216	26 870	13 501	8462	5966	4524	2970	1525	1003	412	147
0.001	10 344	14 403	31 551	141 978	172 011	46 645	13 449	6759	4237	2988	2266	1489	765	504	207	75
0.005	2067	2880	6318	28 473	34 581	9388	2712	1366	858	606	460	303	157	104	44	17
0.01	1032	1440	3164	14 285	17 404	4731	1370	691	435	308	234	155	81	54	23	10
0.05	205	288	641	2938	3670	1009	298	153	98	70	54	37	20	14	7	4
0.10	101	144	327	1525	1966	547	166	87	57	41	32	23	13	10	5	3
0.15	67	96	222	1059	1408	397	123	66	43	32	26	18	11	8	5	4
0.20	50	72	171	830	1138	325	103	56	38	28	23	17	10	8	5	4
0.25	39	58	140	696	985	285	92	51	35	26	21	16	10	8	6	4
0.30	33	49	121	611	892	261	86	48	33	26	21	16	11	9	6	5
0.35	28	42	107	554	836	248	83	47	33	26	21	16	11	9	7	5
0.40	24	38	97	517	803	241	82	48	34	26	22	17	12	10	7	6
0.45	22	34	91	492	790	240	83	49	35	28	23	18	13	11	8	7
0.50	20	32	86	479	793	243	86	51	37	30	25	20	14	12	9	8
0.55	18	30	83	475	812	252	91	55	40	32	27	22	16	14	11	9
0.60	17	28	82	481	849	266	97	59	44	35	30	24	18	16	13	11
0.65	16	28	83	498	908	288	107	66	49	40	34	28	21	18	15	13
0.70	16	28	86	530	997	319	121	75	56	46	40	33	25	22	18	16
0.75	16	29	92	583	1131	366	140	88	67	55	48	39	31	27	22	20
0.80	17	31	103	670	1343	439	171	108	82	68	60	50	39	35	29	26
0.85	18	35	123	826	1708	564	222	143	109	91	80	67	53	47	40	36
0.90	23	45	166	1148	2452	817	327	211	163	137	120	101	81	73	62	56
0.95	37	78	298	2133	4707	1583	641	419	325	273	242	205	165	149	127	116

$\alpha = 0.05$ (two-tailed); $\beta = 0.20$ (power = 80%); control : case ratio = 4 : 1. The sample size listed is the number of subjects needed in the case group. Quadruple this number would be included in the control group.

Table A17 Tabular values of 95% confidence limit factors for estimates of a Poisson-distributed variable.

Observed number on which estimate is based (*n*)	Lower limit factor (*L*)	Upper limit factor (*U*)	Observed number on which estimate is based (*n*)	Lower limit factor (*L*)	Upper limit factor (*U*)	Observed number on which estimate is based (*n*)	Lower limit factor (*L*)	Upper limit factor (*U*)
1	0.0253	5.57	21	0.619	1.53	120	0.833	1.200
2	0.121	3.61	22	0.627	1.51	140	0.844	1.184
3	0.206	2.92	23	0.634	1.50	160	0.854	1.171
4	0.272	2.56	24	0.641	1.49	180	0.862	1.160
5	0.324	2.33	25	0.647	1.48	200	0.868	1.151
6	0.367	2.18	26	0.653	1.47	250	0.882	1.134
7	0.401	2.06	27	0.659	1.46	300	0.892	1.121
8	0.431	1.97	28	0.665	1.45	350	0.899	1.112
9	0.458	1.90	29	0.670	1.44	400	0.906	1.104
10	0.480	1.84	30	0.675	1.43	450	0.911	1.098
11	0.499	1.79	35	0.697	1.39	500	0.915	1.093
12	0.517	1.75	40	0.714	1.36	600	0.922	1.084
13	0.532	1.71	45	0.729	1.34	700	0.928	1.078
14	0.546	1.68	50	0.742	1.32	800	0.932	1.072
15	0.560	1.65	60	0.770	1.30	900	0.936	1.068
16	0.572	1.62	70	0.785	1.27	1000	0.939	1.064
17	0.583	1.60	80	0.798	1.25			
18	0.593	1.58	90	0.809	1.24			
19	0.602	1.56	100	0.818	1.22			
20	0.611	1.54						

Appendix B

Glossary

Accuracy of a measurement is the degree to which the measurement approximates the truth.

Active surveillance is surveillance carried out via a continuous, defined process in a specific population, using one of several approaches. Active surveillance can be medical product based, identifying adverse events in patients taking certain products; setting based, identifying adverse events in certain healthcare settings where patients are likely to present for treatment (e.g., emergency departments); or event based, identifying adverse events likely to be associated with medical products (e.g., acute liver failure).

Actual knowledge, in a legal sense, is defined as literal awareness of a fact. Actual knowledge can be demonstrated by showing that the manufacturer was cognizant of reasonable information suggesting, for example, a particular risk.

Ad hoc studies are studies that require primary data collection.

Adverse drug event, **adverse drug experience, adverse event, or adverse experience** is an untoward outcome that occurs during or following clinical use of a drug. It does not necessarily have a causal relationship with this treatment. It may or may not be preventable.

Adverse drug reaction is an adverse drug event that is judged to be caused by the drug.

Studies of adverse effects examine case reports of adverse drug reactions, attempting to judge subjectively whether the adverse events were indeed caused by the antecedent drug exposure.

Adversomics is the study of vaccine adverse reactions using immunogenomics and systems biology approaches.

Agreement is the degree to which different methods or sources of information give the same answers. Agreement between two sources or methods does not imply that either is valid or reliable.

Analyses of secular trends examine trends in disease events over time and/or across different geographic locations, and correlate them with trends in putative exposures, such as rates of drug utilization. The unit of observation is usually a subgroup of a population, rather than individuals. Also called ecologic studies.

Analytic studies are studies with control groups, such as case–control studies, cohort studies, and randomized clinical trials.

Anticipated beneficial effects of drugs are desirable effects that are presumed to be caused by the drug. They usually represent the reason for prescribing or ingesting the drug.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Anticipated harmful effects of drugs are unwanted effects that could have been predicted on the basis of existing knowledge.

Association is when two events occur together more often than one would expect by chance.

Autocorrelation is where any individual observation is to some extent a function of the previous observation.

Bias is any systematic (rather than random) error in a study.

Biologic inference is the process of generalizing from a statement about an association seen in a population to a causal statement about biologic relationships.

Case–cohort studies are studies that compare cases with a disease to a sample of subjects randomly selected from the parent cohort.

Case–control studies are studies that compare cases with a disease to controls without the disease, looking for differences in antecedent exposures.

Case-crossover studies are studies that compare cases at the time of disease occurrence to different time periods in the same individuals, looking for differences in antecedent exposures.

Case reports are reports of the experience of individual patients. As used in pharmacoepidemiology, a case report usually describes a patient who was exposed to a drug and experienced a particular outcome, usually an adverse event.

Case series are reports of collections of patients, all of whom have a common exposure, examining what their clinical outcomes were. Alternatively, case series can be reports of patients who have a common disease, examining what their antecedent exposures were. No control group is present.

Exposure causes a health event when it truly increases the probability of that event in some individuals. That is, there are at least some individuals who would experience the event given the exposure who would not experience the event absent the exposure.

Changeability is the ability of an instrument to measure a difference in score in patients who have improved or deteriorated.

Channeling bias is a type of selection bias, which occurs when a drug is claimed to be safe and therefore is used in high-risk patients who did not tolerate other drugs for that indication. It is sometimes used synonymously with *confounding by indication.*

Clearance is the proportion of the apparent volume of distribution that is cleared of a drug in a specified time. Its units are volume per time, such as liters per hour. The total body clearance is the sum of clearances by different routes, e.g., renal, hepatic, pulmonary, etc.

Clinical pharmacology is the study of the effects of drugs in humans.

Cohort studies are studies that identify defined populations and follow them forward in time, examining their frequencies (e.g., incidence rate, cumulative incidence) of disease. Cohort studies generally identify and compare exposed patients to unexposed patients or to patients who receive a different exposure.

Combination-triggered drug–drug interaction is, in a potential drug–drug interaction, the scenario in which both the object drug and precipitant drug are initiated simultaneously.

Confidence interval can be conceptualized to represent a range of values within which the true population value lies, with some probability.

Confidentiality is the right of patients to limit the transfer and disclosure of private information.

Confounding by indication can occur when the underlying diagnosis or other clinical features that affect the use of a certain drug are also related to the outcome under study.

Confounding variable, or confounder, is a variable other than the risk factor and outcome variable under study that is related independently both to the risk factor and to the outcome. A confounder can artificially inflate or reduce the magnitude of association between and exposure and outcome.

Constructive knowledge, from a legal perspective, is knowledge that a person did not have, but could have acquired by the exercise of reasonable care.

Construct validity refers to the extent to which results from a given instrument are consistent with those from other measures in a manner consistent with theoretical hypotheses.

Cost is the consumption of a resource that could otherwise be used for another purpose.

Cost–benefit analysis of medical care compares the cost of a medical intervention to its benefit. Both costs and benefits must be measured in the same monetary units (e.g., dollars).

Cost-effectiveness analysis of medical care compares the cost of a medical intervention to its effectiveness. Costs are expressed in monetary units, while effectiveness is determined independently and may be measured in terms of any clinically meaningful unit. Cost-effectiveness analyses usually examine the additional cost per unit of additional effectiveness.

Cost-identification analysis enumerates the costs involved in medical care, ignoring the outcomes that result from that care.

Criterion validity refers to the ability of an instrument to measure what it is supposed to measure, as judged by agreement with a reference (gold) standard.

Cross-sectional studies examine exposures and outcomes in populations at one point in time; they have no time sense.

Data mining is exploratory data analysis for hypothesis generation. As part of a knowledge discovery process, data mining looks to uncover patterns or correlations in the dataset with no or limited presupposition, with the intent of more rigorous testing of any emerging hypothesis tailored to the issue at hand.

Defined daily dose (DDD) is the usual daily maintenance dose for a drug for its main indication in adults.

Descriptive studies are studies that do not have control groups, namely case reports, case series, and analyses of secular trends. They are contrasted with analytic studies.

Detection bias is an error in the results of a study due to a systematic difference between the study groups in the procedures used for ascertainment, diagnosis, or verification of disease.

Differential misclassification occurs when the degree of misclassification of one variable (e.g., drug usage) varies according to the level of another variable (e.g., disease status).

Direct medical costs of medical care are the costs that are incurred in providing the care.

Direct nonmedical costs are nonmedical care costs incurred because of an illness or the need to seek medical care. They can include the cost of transportation to the hospital or physician's office, the cost of special clothing needed because of the illness, and the cost of hotel stays and special housing (e.g., modification of the home to accommodate the ill individual).

Discriminative instruments are those that measure differences among people at a single point in time.

Disease registries are registries characterized by inclusion of subjects based on diagnosis of a common disease or condition.

Drug is any exogenously administered substance that exerts a physiologic effect.

Drug–drug interaction is the phenomenon in which one or more drugs affects the pharmacokinetics and/or pharmacodynamics of one or more other drugs.

Drug utilization, as defined by the World Health Organization (WHO), is the "marketing, distribution, prescription and use of drugs in a society, with special emphasis on the resulting medical, social, and economic consequences."

Drug utilization evaluation (DUE) programs are ongoing structured systems designed to improve drug use by intervening when inappropriate drug use is detected. See also drug utilization review programs.

Drug utilization evaluation studies are *ad hoc* investigations that assess the appropriateness of drug use. They are designed to detect and quantify the frequency of drug use problems.

Drug utilization review programs are ongoing structured systems designed to improve drug use by intervening when inappropriate drug use is detected.

Drug utilization review studies are *ad hoc* investigations that assess the appropriateness of drug use. They are designed to detect and quantify any drug use problems. See also drug utilization evaluation programs.

Drug utilization studies are descriptive studies that quantify the use of a drug. Their objective is to quantify the present state, the developmental trends, and the time course of drug usage at various levels of the healthcare system, whether national, regional, local, or institutional.

Ecologic studies examine trends in disease events over time or across different geographic locations and correlate them with trends in putative exposures, such as rates of drug utilization. The unit of observation is a subgroup of a population, rather than individuals. See also analyses of secular trends.

Effect modification occurs when the magnitude of effect of a drug in causing an outcome differs according to the levels of a variable other than the drug or the outcome (e.g., sex, age group). Effect modification can be assessed on an additive and/or multiplicative scale. See interaction.

Study of drug effectiveness is a study of whether, in the usual clinical setting, a drug in fact achieves the effect intended when prescribing it.

Study of drug *efficacy* is a study of whether, *under ideal conditions*, a drug has the ability to bring about the effect intended when prescribing it.

Study of drug *efficiency* is a study of whether a drug can bring about its desired effect at an acceptable cost.

Enriched or hybrid study designs draw on both primary and secondary data, with some data collected *de novo*, specifically for the purposes of the study and other study-specific data collected via probabilistic or deterministic linkage with other data sources, such as electronic health records, administrative claims and billing data, vital records, and genetic information.

Epidemiology is the study of the distribution and determinants of disease or health-related states in populations.

Evaluative instruments are those designed to measure changes within individuals over time.

Experimental studies are studies in which the investigator controls the therapy that is to be received by each participant, generally using that control to randomly allocate participants among the study groups.

Face validity is a judgment about the validity of an instrument, based on an intuitive assessment of the extent to which an instrument meets a number of criteria including applicability, clarity and simplicity, likelihood of bias, comprehensiveness, and whether redundant items have been included.

Fixed costs are costs that are incurred regardless of the volume of activity.

General causation, from a legal perspective, addresses whether a product is capable of

causing a particular injury in the population of patients like the plaintiff.

Generic quality-of-life instruments aim to cover the complete spectrum of function, disability, and distress of the patient, and are applicable to a variety of populations.

Half-life ($T_{1/2}$) is the time taken for the drug concentration to decline by half. Half-life is a function of both the apparent volume of distribution and clearance of the drug.

Hawthorne effect is when study subjects alter their behavior simply because of their participation in a study, unrelated to the study procedures or intervention.

Health profiles are single instruments that measure multiple different aspects of quality of life.

Health-related quality of life is a multifactorial concept which, from the patient's perspective, represents the end-result of all the physiological, psychological, and social influences of the disease and the therapeutic process. Health-related quality of life may be considered on different levels: overall assessment of wellbeing; several broad domains – physiologic, functional, psychologic, social, and economic status; and subcomponents of each domain – for example pain, sleep, activities of daily living, and sexual function within physical and functional domains.

Human research subject, as defined in US regulations, is "a living individual, about whom an investigator (whether professional or student) conducting research obtains either: (1) data through intervention or interaction with the individual, or (2) identifiable private information" [1].

Hybrid or enriched study designs draw on both primary and secondary data, with some data collected *de novo*, specifically for the purposes of the study, and other study-specific data collected via probabilistic or deterministic linkage with other data sources, such as electronic health records, administrative claims and billing data, vital records, and genetic information.

Hypothesis-generating studies are studies that give rise to new questions about drug effects to be explored further in subsequent analytic studies.

Hypothesis-strengthening studies are studies that reinforce, although do not provide definitive evidence for, existing hypotheses.

Hypothesis-testing studies are studies that evaluate in detail hypotheses raised elsewhere.

Inception cohort design is a cohort study that is restricted to new users of the exposure(s) of interest.

Incidence/prevalence bias, a type of selection bias, may occur in studies when prevalent cases rather than new cases of a condition are selected for a study. A strong association with prevalence may be related to the duration of the disease rather than to its incidence, because prevalence is proportional to both incidence and duration of the disease.

Incidence rate of a disease is a measure of how frequently the disease occurs. Specifically, it is the number of new cases of the disease which develop over a defined time period in a defined population at risk, divided by the number of people in that population at risk.

Indirect costs are costs that do not stem directly from transactions for goods or services, but represent the loss of opportunities to use a valuable resource in alternative ways. They include costs due to morbidity (e.g., time lost from work) and mortality (e.g., premature death leading to removal from the workforce).

Information bias is an error in the results of a study due to a systematic difference between the study groups in the accuracy of the measurements being made of their exposure or outcome.

Instrumental variable is a variable used to adjust for confounding that meets certain specific criteria: it should affect treatment or be associated with treatment choice by sharing a common cause; should be a factor that is as good as randomly assigned, so that it is unrelated to patient characteristics; and should not be related to the outcome other than through its association with treatment.

Intangible costs are those of pain, suffering, and grief.

Interaction, see effect modification.

Interrupted time-series designs include multiple observations of study populations before and after an intervention.

Knowledge, as used in court cases, can be actual or constructive; see those terms.

Large simple trials are randomized trials characterized by large sample sizes, broad entry criteria consistent with the approved medication label, randomization based on equipoise, minimal data requirements, objectively measured endpoints, follow-up that minimizes interventions or interference with normal clinical practice, follow-up of all patients regardless of whether they discontinue randomized medication, and intent-to-treat analysis.

Medication errors are any error in the process of prescribing, transcribing, dispensing, administering, or monitoring a drug, regardless of whether an injury occurred or the potential for injury was present.

Meta-analysis is a systematic, structured review of the literature and formal statistical analysis of a collection of analytic results for the purpose of integrating the findings. Meta-analysis is used to identify sources of variation among study findings and, when appropriate, to provide an overall measure of effect as a summary of those findings.

Microbiome includes the microorganisms, primarily bacteria in the gut, and their genes, harbored within each person.

Misclassification bias is the error resulting from classifying study subjects as exposed when they truly are unexposed, or vice versa. Alternatively, misclassification bias can result from classifying study subjects as diseased when they truly are not diseased, or vice versa.

Molecular pharmacoepidemiology is the study of the manner in which molecular biomarkers alter the clinical effects of medications.

N-of-1 RCT is a randomized controlled trial (RCT) within an individual patient, using repeated assignments to the experimental or control arms.

Near misses are medication errors that have high potential for causing harm but did not, either because they were intercepted prior to reaching a patient, or because the error reached the patient who fortuitously did not have any observable untoward sequelae.

Negative control precipitant drug is, in a study of a potential drug–drug interaction, a drug that is used in similar clinical circumstances as the potential precipitant under study, yet by virtue of the control precipitant's pharmacology is not believed to interact with the study object.

Negative control object drug is, in a study of a potential drug–drug interaction, a drug that is used for similar indications as the object under study, but is not believed to interact pharmacologically with the study precipitant.

Nondifferential misclassification occurs when the misclassification of one variable does not vary by the level of another variable. Nondifferential misclassification usually results in bias toward the null.

Nonexperimental studies are studies in which the investigator does not control the therapy, but observes and evaluates the results of ongoing medical care. The study designs that are used are those that do not involve random allocation, such as case reports, case series,

analyses of secular trends, case–control studies, and cohort studies.

Object drug is, in a drug–drug interaction, the drug(s) whose pharmacokinetics or pharmacodynamics are affected by the other drug(s).

Object-triggered drug–drug interaction is, in a study of a potential drug–drug interaction, the scenario in which the object drug is started in a person already taking the precipitant drug.

Observational studies (or nonexperimental studies) are studies in which the investigator does not control the therapy, but observes and evaluates the results of ongoing medical care. The study designs that are used are those that do not involve randomization, such as case reports, case series, analyses of secular trends, case–control studies, and cohort studies.

Odds ratio is the odds of exposure in the diseased group divided by the odds of exposure in the nondiseased group. When the underlying risk of disease is low (about 10% or lower), it is an unbiased estimator of the relative risk. It is also an unbiased estimate of the rate ratio in a nested or population-based case–control study in which controls are selected at random from the population at risk of disease at the time that the case occurred.

One-group, post-only study design consists of making only one observation on a single group which has already been exposed to a treatment.

Opportunity cost is the value of a resource's next best use, a use that is no longer possible once the resource has been used.

Patient-reported outcomes are any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.

Pharmacodynamics is the study of the relationship between drug level and drug effect. It involves the study of the response of the target tissues in the body to a given concentration of drug.

Pharmacoeconomics is the study of how the price of pharmaceutical products and their economic impact health and the healthcare system.

Pharmacogenetic epidemiology is the study of the effects of genetic determinants of drug response on outcomes in large numbers of people.

Pharmacoepidemiology is the study of the use of and the effects of drugs in large numbers of people. It is also the application of the research methods of clinical epidemiology to the content area of clinical pharmacology, and the primary science underlying the public health practice of drug safety surveillance.

Pharmacogenetics is the study of genetic determinants of responses to drugs. Although it is sometimes used synonymously with pharmacogenomics, it often refers to a candidate-gene approach as opposed to a genome-wide approach.

Pharmacogenomics is the study of genetic determinants of responses to drugs. Although it is sometimes used synonymously with pharmacogenetics, it often refers to a genome-wide approach as opposed to a candidate-gene approach.

Pharmacokinetic compartment is a theoretical space into which drug molecules are said to distribute, and is represented by a given linear component of the log-concentration versus time curve. It is not an actual anatomic or physiologic space, but is sometimes thought of as a tissue or group of tissues that have similar blood flow and drug affinity.

Pharmacokinetics is the study of the relationship between the dose administered of a drug and the concentration achieved in the blood, in the serum, or at the site of action. It includes the study of the processes

of drug absorption, distribution, metabolism, and excretion.

Pharmacology is the study of the effects of drugs in a living system.

Pharmacotherapeutics is the application of the principles of clinical pharmacology to rational prescribing, the conduct of clinical trials, and the assessment of outcomes during real-life clinical practice.

Pharmacovigilance is the identification and evaluation of drug safety signals. More recently, some have also used the term as synonymous with pharmacoepidemiology. The WHO defines *pharmacovigilance* as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other possible drug-related problems [2]. Mann defines pharmacovigilance as "the study of the safety of marketed drugs under the practical conditions of clinical usage in large communities" [3].

Pharmionics is the study of how patients use or misuse prescription drugs in ambulatory care.

Population-based databases or studies refer to whether there is an identifiable population (which is not necessarily based in geography), all of whose medical care would be included in that database, regardless of the provider. This allows one to determine incidence rates of diseases, as well as being more certain that one knows of all medical care that any given patient receives.

Positive control precipitant drug is, in a study of a potential drug–drug interaction, a precipitant drug known to produce an association with an outcome in patients receiving the object drug of interest.

Postmarketing surveillance is the study of drug use and drug effects after release onto the market. This term is sometimes used synonymously with "pharmacoepidemiology," but the latter can be relevant to premarketing studies as well. Conversely, the term "postmarketing surveillance" is sometimes felt to apply only to those studies conducted after drug marketing that systematically screen for adverse drug effects. However, this is a more restricted use of the term than that applied in this book.

Potency refers to the amount of drug that is required to elicit a given response. A more potent drug requires a smaller milligram quantity to exert the same response as a less potent drug, although it is not necessarily more effective.

Potential adverse drug events are medication errors that have high potential for causing harm but did not, either because they were intercepted prior to reaching a patient, or because the error reached the patient who fortuitously did not have any observable untoward sequelae.

Power (statistical power) of a study is the probability of detecting a difference in the study if a difference really exists (either between study groups or between treatment periods).

Pragmatic clinical trials typically fall somewhere between a typical randomized trial and a simple and a large simple trial, where the goal is to introduce one or more pragmatic elements into the design, but with substantial protocol-required follow-up and testing outside of usual care practice.

Precipitant drug is, in a drug–drug interaction, the drug that affects the pharmacokinetics or pharmacodynamics of the other drug(s).

Precipitant-triggered drug–drug interaction is, in a study of a potential drug–drug interaction, the scenario in which the precipitant drug is started in a person already taking the object drug.

Precision is the degree of absence of random error. Precise estimates have narrow confidence intervals.

Precision medicine has been defined by the National Institutes of Health (NIH) in the US

as an "approach to disease prevention and treatment based on people's individual differences in environment, genes and lifestyle" [4].

Pre–post with comparison group design includes a single observation both before and after treatment in a nonrandomly selected group exposed to a treatment (e.g., physicians receiving feedback on specific prescribing practices), as well as simultaneous before-and-after observations of a similar (comparison) group not receiving treatment.

Prescribing errors refer to issues related to underuse, overuse, and misuse of prescribed drugs, all of which contribute to the suboptimal utilization of pharmaceutical therapies.

Prevalence of a disease is a measurement of how common the disease is. Specifically, it is the number of existing cases of the disease in a defined population at a given point in time or over a defined time period, divided by the number of people in that population.

Prevalence study bias is a type of selection bias that may occur in studies when prevalent cases rather than new cases of a condition are selected for a study. A strong association with prevalence may be related to the duration of the disease rather than to its incidence, because prevalence is proportional to both incidence and duration of the disease.

Privacy, in the setting of research, refers to each individual's right to be free from unwanted inspection of, or access to, personal information by unauthorized persons.

Procedure registries are registries characterized by inclusion of subjects based on receipt of specific services, such as procedures, or based on hospitalizations.

Product registries are registries characterized by inclusion of subjects based on use of a specific product (drug or device) or related products in a given therapeutic area.

Propensity scores are an approach to controlling for confounding that uses mathematical modeling to predict exposure based on observed variables, and uses the predicted probability of exposure as the basis for matching or adjustment.

Prospective drug utilization review is designed to detect drug-therapy problems before an individual patient receives the drug.

Prospective studies are studies performed simultaneously with the events under study; namely, where patient outcomes have not yet occurred at the outset of the study.

Proteomics is, within the context of pharmacoepidemiology, the study of how proteins are responsible for variability in medication response.

Protopathic bias is interpreting a factor to be a result of an exposure when it is in fact a determinant of the exposure, and can occur when an early sign of the disease under study led to the prescription of the drug under study.

Publication bias occurs when publication of a study's results is related to the study's findings, such that study results are not published or publication is delayed because of the results.

P value is the probability that a difference as large as or larger than the one observed in the study could have occurred purely by chance if no association truly existed.

Qualitative drug utilization studies are studies that assess the *appropriateness* of drug use.

Quality of life is the description of aspects (domains) of physical, social, and emotional health that are relevant and important to the patient.

Quantitative drug utilization studies are descriptive studies of *frequency* of drug use.

Random allocation is the assignment of subjects who are enrolled in a study into study groups in a manner determined by chance.

Random error is error due to chance.

Random selection is the selection of subjects for a study from among those eligible in a manner determined by chance.

Randomized clinical trials are studies in which the investigator randomly assigns patients to different therapies, one of which may be a control therapy.

Recall bias is an error in the results of a study due to a systematic difference between the study groups in the accuracy or completeness of their memory of their past exposures or health events.

Referral bias is error in the results of a study that occurs when the reasons for referring a patient for medical care are related to the exposure status, e.g., when the use of the drug contributes to the diagnostic process.

Registries are organized systems that use observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. Registries can be thought of as both the process for collecting data from which studies are derived, as well as referring to the actual database.

Regression to the mean is the tendency for observations on populations selected on the basis of an abnormality to approach normality on subsequent observations.

Relative rate is the ratio of the incidence rate of an outcome in the exposed group to the incidence rate of the outcome in the unexposed group. It is synonymous with the terms *rate ratio* and *incidence rate ratio*.

Relative risk is the ratio of the cumulative incidence of an outcome in the exposed group to the cumulative incidence of the outcome in the unexposed group. It is synonymous with the term *cumulative incidence ratio*.

Reliability is the degree to which the results obtained by a measurement procedure can be replicated. The measurement of reliability does not require a gold standard, since it assesses only the concordance between two or more measures.

Reporting rate in a spontaneous reporting system is the number of reported cases of an adverse event of interest divided by some measure of the suspect drug's utilization, usually the number of dispensed prescriptions. This is perhaps better referred to as a *rate of reported cases*.

Reproducibility is the ability of an instrument to obtain more or less the same scores on repeated measurement of patients who have not changed.

Research, as defined in US regulations, is any activity designed to "develop or contribute to generalizable knowledge" [5].

Research subject is "a living individual, about whom an investigator (whether professional or student) conducting research obtains either: 1) data through intervention or interaction with the individual, or 2) identifiable private information" [6].

Responsiveness is an instrument's ability to detect change.

Retrospective drug utilization review compares past drug use against predetermined criteria to identify aberrant prescribing patterns or patient-specific deviations from explicit criteria.

Retrospective studies are studies conducted after the events under study have occurred. Both exposure and outcome have already occurred at the outset of the study.

Risk is the cumulative probability that something will happen.

Risk evaluation and mitigation strategy (REMS) is a pharmacovigilance assessment plan in the US, approved by regulators in advance of implementation, to ensure that the benefits of a drug or biologic product outweigh its risks.

For a **risk management plan** in the EU, pharmacovigilance legislation explicitly requires the active monitoring of the outcome of risk minimization activities it contains, placing the obligation on manufacturers and regulatory authorities for this activity.

Judgment about **safety** is a personal and/or social judgment about the degree to which a given risk is acceptable.

Safety signal is a concern about an excess of adverse events compared to what is expected to be associated with use of a product (drug or device).

Sample distortion bias is another name for selection bias.

Scientific inference is the process of generalizing from a statement about a population, which is an association, to a causal statement about scientific theory.

Selection bias is error in a study that is due to systematic differences in characteristics between those who are selected for the study and those who are not.

Self-controlled designs are studies that include only persons who experienced the outcome, using each person as their own control, and include self-controlled case series and case-crossover designs.

Self-controlled case series (SCCS) design is a self-controlled design that is analogous to the cohort design. It includes only individuals who experienced the outcome, and examines the rate of the outcome during exposed vs. unexposed periods within those individuals.

Sensibility is a judgment about the validity of an instrument, based on an intuitive assessment of the extent to which an instrument meets a number of criteria including applicability, clarity and simplicity, likelihood of bias, comprehensiveness, and whether redundant items have been included.

Sensitivity is the proportion of persons who truly have a characteristic, who are correctly classified by a diagnostic test as having it.

Sensitivity analysis is a set of procedures in which the results of a study are recalculated using alternate values for some of the study's variables, in order to test the sensitivity of the conclusions to altered specifications.

Serious adverse experience is any adverse experience occurring at any dose that results in any of the following outcomes: death, a life-threatening adverse experience, inpatient hospitalization or prolongation of existing hospitalization, a persistent or significant disability/incapacity, or congenital anomaly/birth defect.

Service registries are registries characterized by inclusion of subjects based on receipt of specific services, such as procedures, or based on hospitalizations.

Signal is a hypothesis that calls for further work to be performed to evaluate that hypothesis.

Signal detection is the process of looking for or identifying signals from any source.

Signal generation, sometimes referred to as data mining, is an approach that uses statistical methods to identify a safety signal. No particular medical product exposure or adverse outcome is prespecified.

Signal refinement is a process by which an identified safety signal is further evaluated to determine whether evidence exists to support a relationship between the exposure and the outcome.

Specific causation, from a legal perspective, addresses whether the product in question actually caused an alleged injury in the individual plaintiff.

Specificity is the proportion of persons who truly do *not* have a characteristic, who are correctly classified by a diagnostic test as not having it.

Specific quality-of-life instruments are focused on disease or treatment issues specifically relevant to the question at hand.

Spontaneous reporting systems are maintained by regulatory bodies throughout

the world and collect unsolicited clinical observations that originate outside of a formal study.

Statistical inference is the process of generalizing from a sample of study subjects to the entire population from which those subjects are theoretically drawn.

Statistical interaction, see effect modification.

Statistically significant difference is a difference between two study groups that is unlikely to have occurred purely by chance.

Steady state, within pharmacokinetics, is the situation when the amount of drug being administered equals the amount of drug being eliminated from the body.

Systematic error is any error in study results other than that due to random variation.

Therapeutic ratio is the ratio of the drug concentration that produces toxicity to the concentration that produces the desired therapeutic effect.

Therapeutics is the application of the principles of clinical pharmacology to rational prescribing, the conduct of clinical trials, and the assessment of outcomes during real-life clinical practice.

Type A adverse reactions are those that are the result of an exaggerated but otherwise predictable pharmacologic effect of the drug. They tend to be common and dose related.

Type B adverse reactions are those that are aberrant effects of the drug. They tend to be uncommon, not dose related, and unpredictable.

Type I statistical error is concluding there is an association when in fact one does not exist, i.e., erroneously rejecting the null hypothesis.

Type II statistical error is concluding there is no association when in fact one does exist, i.e., erroneously accepting the null hypothesis.

Unanticipated beneficial effects of drugs are desirable effects that could not have been predicted on the basis of existing knowledge.

Unanticipated harmful effects of drugs are unwanted effects that could not have been predicted on the basis of existing knowledge.

Uncontrolled studies refer to studies without a comparison group.

Unexpected adverse experience means any adverse experience that is not listed in the current labeling for the product. This includes an event that may be symptomatically and pathophysiologically related to an event listed in the labeling, but differs from the event because of greater severity or specificity.

Utility measures of quality of life are measured holistically as a single number along a continuum, e.g., from death (0.0) to full health (1.0). The key element of a utility instrument is that it is preference based.

Vaccinovigilance is all methods of assessment and prevention of adverse events following immunizations.

Validity is the degree to which an assessment (e.g., questionnaire or other instrument) measures what it purports to measure.

Variable costs are costs that increase with increasing volume of activity.

Apparent **volume of distribution** (V_D) is the apparent volume that a drug is distributed in after complete absorption. It is usually calculated from the theoretical plasma concentration at a time when all of the drug was assumed to be present in the body and uniformly distributed. This is calculated from back extrapolation to time zero of the plasma concentration time curve after intravenous administration.

Voluntariness is the concept in research ethics that investigators must tell subjects that participation in the research study is voluntary, and that subjects have the right to discontinue participation at any time.

References

1 Title 45 US Code of Federal Regulations Part 46.102 (f).
2 World Heath Organization. *Safety Monitoring of Medicinal Products: The Importance of Pharmacovigilance*. Geneva, WHO, 2002.
3 Mann RD, Andrews EB, eds., *Pharmacovigilance*. Chichester: John Wiley & Sons, 2002.
4 National Institutes of Health. The future of heath begins with you. https://www.nih.gov/precision-medicine-initiative-cohort-program (accessed May 2019).
5 Title 45 US Code of Federal Regulations Part 46.102 (d).
6 US Code of Federal Regulations 46.102 f.

Index

a

abatacept 569
ABCB *see* ATP-binding cassette transporter proteins
absolute risk 1012, 1016
absorption 29
abuse of medications 993
academia
  drug approval process 84–5
  economic assessment of medication-related issues 91
  funding 1117
  future 92–3, 1111–12, 1112–18
  interventional pharmacoepidemiology 90–1
  logistical advances 1115–16
  medical centers 153–4
  methodologic advances 1112–13
  new content areas of interest 1113–15
  personnel 1117–18
  pharmacoepidemiology role in 83–93
  policy analysis 89–90
  prescribing practices 85–6
  quality and outcomes of drug use 88–9
  scientific developments 1112–16
academic detailing 90, 420–1
Academic Medical Center 91, 92
accelerated approval 129, 807
Accelerated Development of Vaccine Benefit risk Collaboration in Europe *see* ADVANCE
acceptability
  metaanalyses 930
  risk evaluation studies 1010
  see also risk tolerance
acceptable mortality reporting (AMR) 256
acceptance, definition 991
accuracy of data
  diagnosis 973–4
  recall 960–8, 969
  validity 959–62, 969, 965–6, 970, 972–4
ACE *see* angiotensin converting enzyme inhibitors
acetaminophen 586, 601–2
activated protein C (APC) 301–2
Active Permanent Workshop of Imputologists (APWI) 729, 736
active surveillance 782
  see also Sentinel Initiative
actual knowledge 143–4
acute myeloid leukemia (AML) 299–300
acute myocardial infarction (AMI) 1085
ad hoc epidemiologic studies 461, 1112
adalimumab 569
ADAPTABLE Aspirin Study 807
ADAs *see* antidrug antibodies
addiction, definition 702
Addiction Severity Index-Multimedia Version (ASI-MV®) surveillance network 715
additional risk minimization measures (aRMMs) 100
ADE *see* adverse events/adverse drug reactions
Adenomatous Polyp Prevention on Vioxx (APPROVe) study 680
ADHD *see* attention-deficit/hyperactivity disorder
adherence, definition 86
  see also patient compliance/adherence
ADME studies *see* pharmacokinetics
administrative claims 970, 972, 973–4
administrative data
  Cerner Health Facts Database 292
  characteristics 212
  claims data 511–13
  disadvantage 208
  medical devices 511–13
  medication errors 1047–8
  pediatric 292–3
  validity 951, 969–79
  Veterans Health Administration 258
administrative databases 206–7
ADR *see* adverse events/adverse drug reactions
ADVANCE 464, 621
adverse event reporting systems (AERS) 179, 709
adverse events following immunizations (AEFI) 437, 438, 440, 441–2, 443, 456–466, 469, 471–3, 690
adverse events of interest (AEIs) 337, 565–6
adverse events/adverse drug reactions (AE/ADR) 3, 4–5
  bioethics 786
  biologics 565–6
  case reports 725–740
  causation 725–740
  data mining 677–8
  definitions and concepts 5
  drug utilization 381
  drug–drug interactions 1031

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

adverse events/adverse drug reactions (AE/ADR) (*cont'd*)
event monitoring 312–19, 325–9, 331–2, 334, 335
inherent susceptibility to 1041
intellectual development 10–13
legal factors 74–5
legislation and regulations 6–10
legislative instruments 156
medication errors 1047, 1049–54
metaanalyses 899, 922–3
molecular pharmacoepidemiology 752–3
optimization of therapy 4
pharmacovigilance reporting systems 169
postmarketing studies 73
prescribing practices 424, 425
reporting system 777, 778
risk evaluation studies 1011, 1013, 1015
risk management 582, 589–90
spontaneous reporting 169–96
vaccines 437, 438, 440, 441, 443–53
see also birth defects
AED *see* antiepileptic drugs
AEFI *see* adverse events following immunizations
AERS *see* adverse event reporting systems
Aetna 360
age effects
recall accuracy 962
validity of data 966
see also elderly populations
Agency for Health Care Policy and Research 639
Agency for Healthcare Research and Quality (AHRQ) 12, 349
comparative effectiveness research 639, 650, 657, 659, 660, 662
drugs of abuse 707
future developments 1117, 1119
agranulocytosis 1084–5
agreement 949
AHFS *see* American Hospital Formulary Service
AHRQ *see* Agency for Healthcare Research and Quality
AIDS Clinical Trials Group (ACTG) adherence questionnaire 995
AJRR *see* American Joint Replacement Registry (AJRR)
alemzumab 609–11
alert fatigue 1031
algorithm methods 734–5
allergies 1050
ALLHAT study 762–3, 765
alloyed (tarnished) gold standard 949
α-chymotrypsin and anaphylaxis 195
α-adducin gene 753
α errors *see* Type I/II errors
AMD *see* age-related macular degeneration
American Hospital Formulary Service (AHFS) 226
American Joint Replacement Registry (AJRR) 503
American Medical Association (AMA)
Committee on Pharmacy and Chemistry 773
Council on Pharmacy and Chemistry 6
American Recovery and Reinvestment Act (ARRA Stimulus) 639
AMI *see* acute myocardial infarction
AML *see* acute myeloid leukemia
amoxicillin 1036, 1037, 1038
analysis of secular trends 50, 51–2
Anatomical Therapeutic Chemical (ATC) Classification System 626
codes 254
drug utilization 386, 390–1
anchor-based methods1023, 1071
angiotensin-converting enzyme (ACE) inhibitors 392, 658, 678
animal models 452, 588, 589
antibiotics
adverse events 299, 1053
case–crossover design 1037–8
children 538
chronic obstructive pulmonary disease 300–1, 1085
drug utilization studies 376–7, 395
group education 421
hepatic dysfunction 36
in-hospital databases 254–5
inpatient databases 300–1
overuse 412
pregnancy 544
sepsis and 302
uses 421, 815
anticipated beneficial effects 813–14
anticipated harmful effects 813
anticoagulants 823–4, 825
antidepressants 924–5, 930
antidrug antibodies (ADAs) 566, 568, 569
antiepileptic drugs (AED) 330–1, 931
Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) 762–3,765
antiretrovirals 129, 156, 396, 793, 998, 1000, 102
anti-TNF (tumor necrosis factor) drugs 562, 564, 565, 568, 568
APC *see* activated protein C
aplastic anemia 6, 186–8, 370, 586, 813
apparent volume of distribution 29
APWI *see* Active Permanent Workshop of Imputologists
Arianna database 253, 254, 262, 264
ARITMO *see* Arrhythmogenic Potential of Drugs
Armed Forces Institute of Pathology 728
Arrhythmogenic Potential of Drugs (ARITMO) 266, 621
artifactual associations 46
ASCQ-Me^{R} 1073
Asian Pharmacoepidemiology Network (AsPEN) 618, 630, 779, 934
Code of Conduct for Collaborative Vaccine Studies 785
aspirin 30, 91, 264, 380, 425, 438, 631, 800, 802, 807, 814, 909, 1037
associations
criteria for causation 47–50
definitions and types 46–7
validity of data 963, 969–70
asthma 7, 17, 90, 118, 186, 332, 412, 1083, 1084, 1086, 1088–91
ATC *see* Anatomical Therapeutic Chemical Classification Scheme
ATP-binding cassette transporter proteins (ABCB) 751–2
atrial fibrillation 873, 876, 878–9, 880–1, 884
attention-deficit/hyperactivity disorder (ADHD) 653
database size 233
treatment 193
audit systems 422–3
automated healthcare databases 543–4
see also electronic databases
automated signal generation 333

b

Bar Code Medication Administration (BCMA) database 258
barcoding 1055

Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria *see* BIFAP
Bayesian Confidence Propagation Neural Network (BCPNN) 185, 194
Bayesian methods
  causation 735–8, 739
  data-mining algorithms 709
  medical devices 506
  metaanalyses 914–15, 917. 925, 927, 930
Bayesian Shrinkage 682
Bayh–Dole Act 838
BCMA *see* Bar Code Medication Administration
BCPA *see* Best Pharmaceuticals for Children Act
BCPNN *see* Bayesian Confidence Propagation Neural Network
BED-FRAME *see* Benefit-Risk Evaluation of Diagnostics: A Framework
before and after studies 332–3
behavioral economics 841–3
Bendectin® 7, 145
beneficial drug effects 813–28
  anticoagulants for MI death prevention 825
  anticoagulants for VTE 823–4
  clinical problems 814–16
  confounding by indication 816–18, 822, 824–5,826, 828
  cost-effectiveness studies 826–7
  current solutions 818–27
  definitions 813–14
  estrogens for osteoporosis 822–3
  expected 366
  future 827–8
  generic versus brand name drugs 825–6
  lidocaine for MI death prevention 824–5
  methodologic problems 816–18
  vaccines 827
benefit–risk assessments 867–90
  clinical problems 868–9
  communicating 870
  current solutions 870–85
  data for 869–70, 874–6
  decision context 871
  definition 868
  frameworks 870–1, 876–7
  future 886–90
  initiatives, guidances, and partnerships 886
  integrated assessment 879–82
  integrating benefits and risks 870
  methodologic problems 869–70
  outcome identification and definition 872–4
  patient perspective 868–9
  systematic approach to 868
  value judgments and patient preferences 877–9
  value tree 872, 873
  weighting and patient perspective 882–5
Benefit-Risk Evaluation of Diagnostics: A Framework (BED-FRAME) 890
benefit–risk profiles
  biopharmaceutical industry 99, 115
  legislative instruments 142, 147, 149
  medical devices 496, 498
  regulatory agencies 134, 135
  risk evaluation studies 1014, 1015
  risk management 581, 582, 588, 589
  vaccines 439
benfluorex scandal 777–8
benoxaprofen 7
benzodiazepine 714–15
Best Pharmaceuticals for Children Act (BCPA) (2002) 530
β(2)-adrenergic receptors 752
beta-agonists 1088–91
beta-blockers
  beneficial drug effects 818
  metaanalyses 926–8
  prescribing practices 86
  risk tolerance 78
beta errors *see* Type I/II errors
BfArM *see* Federal Institute for Drugs and Medical Devices
biases
  children 525
  clinical epidemiology 46
  event monitoring 329–30
  metaanalyses 898–9, 899–900, 902–5, 905–6, 909, 901, 911–12, 913, 914–15, 916–17, 918–20
  prescribing practices 419
  publication 903–5, 918–20, 930, 931
  vaccines 445–6
  validity of data 950, 951, 952, 953–5, 959, 965, 969
  see also confounding data
BIFAP 253, 254, 262, 264
Big Data 115–19, 787
Bill and Melinda Gates Foundation 838
billing data 345,347
  administrative 512
  Medicaid 11
  pharmaceutical 387, 571
Bilski v. Kappos 157
BioBank 349, 350
bioethics *see* ethics
biological inference 45
biological plausibility 47, 761
biologics 561–74
  adverse events 565–6
  biosimilar naming conventions 571
  clinical problems 562–72
  comparative effectiveness 566–8
  confounding by indication 571–2
  future 572–4
  identification 569–71
  measuring treatment discontinuation 572
  methodological problems 569–74
  neoplasia 564–5
  new data sources and linked data 573–4
  new therapies 572–3
  paradoxical reactions 565
  serious infections 562–3
  switching and restarting 568–9
  value-based care 573–4
Biologics License Applications (BLA) 605
biopharmaceutical industry 98–119
  Big Data 115–19
  causation 730–1
  data sources 102–3
  drug safety evaluations 101–12
  future developments 115–19, 1111, 1118–19
  legislative instruments 140–59
  postapproval safety studies 106–8
  preapproval safety studies 103–6
  risk management 99–100
  risk mitigation interventions 99, 105, 112–15
  special populations 108–12
  study design 101–2
biosimilar naming conventions 571
Biosimilar User Fee Act (BsUFA) (2012) 11
biosimilars 569
Bioventus Exogen® device registry 349

birth defects 534, 535
acceptability of drug risk 78, 79
bias 46
biological plausibility 528
legislative instruments 145
primary data collection 343, 348
product liability law 151
risk management 602
risk tolerance 78
see also thalidomide
birth registries 117
BLA *see* Biologics License Applications
black box warning 714
black triangle program 181
blinding (masking) 343, 627, 656
randomized controlled trials 795–7
body composition 32–3, 34, 35
Bonferroni correction 764
Boston Collaborative Drug Surveillance Program 6, 11, 360, 774
Boston University Fever Study (BUFS) 802–4
Bradford Hill criteria 146, 681, 713, 728
brand name drugs 825–6
bretylium 74
Brief Medication Questionnaire 995
Brighton Collaboration 444, 447
Brilinta® (ticagrelor) 194, 600
Brintellix® (vortioxetine) 194, 600
British National Formulary 181, 390
bromfenac sodium 586–7
BUFS *see* Boston University Fever Study
bupropion 611–12
Burroughs Wellcome Foundation 1117–18

C

CADEUS cohort 968
calendar-time cohorts 1081–2
Canadian Institute for Health Information (CIHI) 226
Canadian Network for Observational Drug Effect Studies (CNODES) 230, 617, 624, 632, 934
Canadian Pharmacogenomics Network for Drug Safety 545
Cancer and Steroid Hormone (CASH) study 347
Cancer Drugs Fund 640
Cancer Moonshot concept 843
cancer registries 227, 263, 346–7, 350, 512, 762, 806
Cancer Research Network 618
cancer
lung 46–50, 52, 148
metastatic breast cancer 344, 348, 753
cancer-targeted therapies 756–7
candidate gene approaches 761–2
capture–recapture methods 458
cardiac structure and function 32, 36
Cardiovascular Research Network 618
CAR-T therapy 838
case–cohort studies 1079, 1081, 1085–6
case–control studies
ad hoc 363, 364, 366, 367
beneficial drug effects 822, 823, 824, 825
causation 737
children 544–5
clinical epidemiology 52, 53–4, 55, 56–7
confounding data 1079, 1081, 1082–5
drug safety evaluations 102
event monitoring 328, 330
molecular pharmacoepidemiology 755, 758–61, 762
pregnancy 524–5, 540
registry based 363, 366
sample size 60, 62, 65–6, 68–9
vaccines 445, 468
validity of data 953
case–control surveillance 360, 364, 365, 367, 369, 370
case-crossover study design 101–2, 103, 1088–90
case report forms 346
case reports
algorithm/criterial methods 734–5
uses of causality assessment 729
causation 725–40
clinical epidemiology 50–1, 55, 56–7
clinical problems 725–7
comparison of methods 738–9
current solutions 733–9
future developments 739–40
historical perspectives 727–32
methodologic problems 732–3
pharmaceutical manufacturers 730–1
pharmacoeconomics 845, 846
probabilistic methods 735–8
publishers of AE reports 732
regulatory factors 731–2
scoring of individual judgments 734–5
spontaneous reporting 175–7, 177–9, 183, 184, 185–6
unstructured clinical judgments/global introspection 728, 733–4
verbal judgments 734
case series
clinical epidemiology 51, 55, 56
event monitoring 328
sample size 66–8
spontaneous reporting systems 184–6, 188
case-time-control studies 1090–1
CASH *see* Cancer and Steroid Hormone study
causation
algorithm/criterial methods 734–5
case reports 725–40
clinical epidemiology 45
clinical problems 725–7
comparison of methods 738–9
criteria for an association 47–50
current solutions 733–9
data mining and 681–3
definition 45
future developments 739–40
historical perspectives 727–32
legislative instruments 145–6
methodologic problems 732–3
molecular pharmacoepidemiology 753
pharmaceutical manufacturers 730–1
probabilistic methods 735–8
publishers of AE reports 732
regulatory factors 731–2
scoring of individual judgments 734–5
spontaneous reporting 171, 181–2, 184, 186, 190, 191
unstructured clinical judgments/global introspection 728, 733–4
uses of causality assessment 729
vaccines 444, 452–4
verbal judgments 734
causation of damages 145–6
general causation 145–6
specific causation 146
reasonable certainty 146
CBA *see* cost–benefit analysis
CDER *see* Center for Drug Evaluation and Research
CDM *see* common data model

CDRH *see* Center for Devices and Radiological Health
CDSS *see* clinical decision support systems
cefepime 299–300
Cegedim 252
celecoxib 75, 313, 331, 334
Center for Devices and Radiological Health (CDRH) 878
Center for Drug Evaluation and Research (CDER) 349, 531
Centers for Disease Control and Prevention (CDC) 92, 545
  future developments 1119
  Office of Public Health Genomics 756
  vaccines 448–9, 451,455, 462
Centers for Education and Research on Therapeutics (CERTs) 11–12
  future developments 1117
Centers for Medicare and Medicaid Services (CMS) 218, 222–3, 298
  see also Medicaid; Medicare
CER *see* comparative effectiveness research
cerivastatin (Baycol®) 141, 144, 175
Cerner Health Facts (HF) 290, 292
CERT *see* Centers for Education and Research on Therapeutics
certolizumab 570
CESR *see* Comparative Effectiveness and Safety Research
CHCA *see* Child Health Corporation of America
chemotherapy
  children 299–300, 1049
  comparative effectiveness research 641
  patient-reported outcomes 1064
  risk management 586
  weight-based dosing 1049–50
CHF *see* Commissie Farmaceutische Hulp
Child Health Corporation of America (CHCA) 292
children 524–47
  automated healthcare databases 543–4
  bias in samples 525
  biology and epidemiology 525–6
  case–control studies 545
  clinical practice 530–1
  confounding 538
  currently available solutions 541–6
  drug safety evaluations 111–12
  ethics and regulations 528–9, 529–30
  exposure ascertainment, timing, and misclassification 532–3, 534
  future 546–7
  growth and development 528
  medication errors 1049–50, 1051
  methodological problems 531–41
  newer designs 546
  outcome and ascertainment 534–5, 535–6
  pharmacokinetics 34–5
  population definition 531–2
  prospective cohort 541
  randomized controlled trials 802–4
  registries 542
  sample size requirements and challenges 532
  selection bias 540–1
  self-controlled designs 546
  treatment responses and patterns 527–8
Children's Hospital Association (CHA) 292
Children's Health Insurance Program (CHIP) 223
Children's Oncology Group 299–300
chloramphenicol 6, 397, 813
CHM *see* Commission on Human Medicines
CHMP *see* Committee for Medicinal Products for Human Use
cholestatic jaundice 64
chronic care model
  self-management support 1065
chronic conditions
  causality 729
  children 544
  comparative effectiveness research 600
  confounding 1080, 1090
  event monitoring 330
  patient compliance/adherence 998
  pharmacoeconomics 818
  pregnancy 537
  prescribing practices 85, 87
  risk management 604
  validity of drugs 965
chronic kidney disease (CKD) 192
chronic obstructive pulmonary disease (COPD) 300–1, 332, 874, 1085
Chronic Respiratory Questionnaire (CRQ) 1071
CIHI *see* Canadian Institute for Health Information
CIHR *see* Canadian Institute for Health Research
cimetidine 4, 51, 66, 72, 75, 776
CIOMS *see* Council for International Organizations of Medical Sciences
CISA *see* Clinical Immunization Safety Assessment
cisapride 8, 144, 585
CKD *see* chronic kidney disease
claims data
  electronic databases 206–7
  medical devices 511–13
  medication errors 1047–8
  see also commercial insurance databases; government claims databases
Clarification of Optimal Anticoagulation through Genetics (COAG) trial 755–6
clearance 34
Clinical and Translational Science Awards (CTSA) 1118
clinical benefits 75
clinical data 263–4
Clinical Data Research Networks 621
clinical decision support systems (CDSS) 1055
clinical economics 845, 846–7
clinical epidemiology 44–57
  analyses of secular trends 51–2, 55, 57
  case–control studies 52, 53–4
  case reports 50–1, 55, 56–7
  case series 51, 55, 56
  cohort studies 52–3, 53–4
  criteria for causation 47–50
  errors 46–7
  randomized controlled trials 44, 50
  scientific method 44–5
  study design 50–5
Clinical Immunization Safety Assessment (CISA) 442, 451–2
clinical pharmacology 27–41
  basics of 28–9
  definition 4
  model-informed drug development 40–1
  pharmacodynamic 37–40
  pharmacoepidemiology 28
  pharmacogenomics 40
  pharmacokinetics 29–32
  risk management 588
  special populations 32–7

Clinical Practice Research Datalink® (CPRD®) 252–7, 263, 328, 363, 542, 806
Clinical Research Networks (CRNs)
  event monitoring 310
  medical devices 509–11
clinical trials 10, 13, 14–15
  beneficial drug effects 823–4
  causation 737
  comparative effectiveness research 652, 655
  costs 71
  ethics 785
  future developments 1115
  hypothesis generating studies 75–6
  legislative instruments 143, 144–5, 1492, 153
  marketing 73–4
  medical devices 501–2
  metaanalyses 898, 899–901, 903–4, 923–4, 926, 928, 931
  patient compliance/adherence 1001
  pharmacogenomics 40
  pharmacokinetics 34, 40
  risk management 588, 589, 599
  spontaneous reporting 170
  vaccines 441–2, 460–1
  see also randomized controlled trials
Clinical Utility Gene Cards (CUGCs) 756
clinical-reported outcomes 1063
clioquinol 7, 775
clomifene citrate 195
clopidogrel 86, 89, 91, 1017, 1030, 1036, 1040
cluster randomization 417, 655, 807, 917
cluster trials 807
CMS *see* Centers for Medicare and Medicaid Services
CNODES *see* Canadian Network for Observational Drug Effect Studies
CNV *see* copy number variants
COAG *see* Clarification of Optimal Anticoagulation through Genetics (COAG) trial
Cochrane Collaboration
  metaanalyses 897, 898, 902, 905, 907–9, 911, 924, 930
  prescribing practices 412, 414
coding errors 626
coding systems
  adverse reactions 183
  drug 390
  drug validation 971–4, 979
  electronic health record databases 254
  encounter database 220, 225, 226, 228, 232
  epidemiology 103
COFEPRIS (Federal Commission for Protection against Health Risks) (Mexico) 100
coherence 47
cohort studies 52–3, 53–4, 55
  ad hoc 365, 366, 367, 368
  beneficial drug effects 822, 834
  causation 737
  confounding data 1079–93, 1099–101
  drug abuse 710, 711
  event monitoring 309, 312, 316, 318–19, 321, 325, 327–8
  medical devices 505
  sample size 60, 61–4, 68–9
  structures of cohorts 1081–2
  vaccines 445, 467–8
  validity of data 971
collaboration
  legislative instruments 152
  medical devices 516
  regulatory agencies 134–6
combination triggered concomitancy 1040
Comment k 142
commercial insurance databases 214, 217, 218–22
Commissie Farmaceutische Hulp (CHF) 640, 641
Commission on Human Medicines (CHM) 307
Committee for Human Medicinal Product (CHMP) 15
  marketing authorization 777
  risk management 594, 595
  risk minimization 608, 611
Committee for Pharmaceutical Aid (Netherlands) 640
Committee on Safety of Drugs (CSD) (UK) 307, 308
Committee on Safety of Medicines (CSM) 307, 308
  Working Party on Adverse Reactions 308
common data model (CDM) 687–8
  definition 618
  necessity of having 627–8
  with or without common protocol 624–6
common protocol 618, 626–7
Common Terminology Criteria for Adverse Events (CTCAE) scale 877
communicating risks 585, 589–90, 1011–13
community pharmacies 212, 1021–2
co-morbidities 74, 104, 107, 130, 1052, 1053
  drugs of abuse 704, 708
  event monitoring 310
  randomized controlled trials 804, 809
comparative effectiveness research (CER) 12, 92, 349, 639–63
  African countries 642
  biologics 566–8
  China, India, and South Korea 642
  clinical practice 662–3
  clinical problems 646–8
  commercial insurance databases 214, 217, 218–22
  confounding by indication 650–2
  cost containment 659
  cost-effectiveness analysis 659
  current solutions 659–60
  data sources, record linkage, and multidatabase studies 652–5
  definitions, key components and goals 642–4
  efficacy vs. effectiveness 642
  Europe 640–1
  evidence dissemination 658–9
  evidence generation 650–7
  evidence synthesis 648–50
  experimental studies 655–7
  funding 660–1
  future developments 660–3
  healthy user/candidate bias 652
  Hong Kong 641–2
  human capital development 661–3
  key attributes 646–8
  medical devices 502, 512
  methodologic problems 648–59
  nonexperimental studies 650–2
  observational 649–50, 656–7
  organizational approaches 656–7
  published guides 657
  scope of 646
  systematic reviews and metaanalysis 648–9
  US 639–40
Comparative Effectiveness Research through Collaborative Electronic Reporting (CER2) Consortium 544

comparative studies
confounding data 1094–5, 1100
medical devices 504
metaanalyses 921–2
validity of data 974
COMPARE project 682
COMPASS *see* Computerized Online Medicaid Analysis and Surveillance System
competent regulatory authorities 592, 593, 594
complementary medicines 1114
complete paid-claims files of drug benefit programs 87
completeness of data
validity 970–1, 972, 974, 979
compliance, definition 86
see also patient compliance/adherence
composite indices 1070
computerized decision support 424
Computerized Online Medicaid Analysis and Surveillance System (COMPASS) 11, 826
computerized physician order entry (CPOE) 418, 658, 1048, 1055–6
concentration–effect relationship 37–8
concomitant administration 1040
concordance 949
concurrent comparison groups 414
confidence intervals
clinical epidemiology 54
metaanalyses 913–14
sample size 60, 67
confidentiality *see* privacy and confidentiality
confirmatory data analysis (CDA) 685
confounders, important 367
confounding by disease 537
confounding by indication
beneficial drug effects 816–18, 820–1, 822, 824–5, 826, 828
biologics 571–2
children 537
comparative effectiveness research 650–2
data mining 684
drug–drug interaction 1031–5
drug safety evaluations 101
pregnancy 537
randomized controlled trials 798
regulatory agencies 127
confounding data 1078–1101
balancing patient risk factors 1091
case–cohort studies 1078, 1085–6
case–control studies 1079, 1081, 1082–5
clinical epidemiology 46
clinical problems 1078–9
children 538
cohort studies 1078, 1079, 1081–93, 1099–100
comparative effectiveness research 650–2
comparative studies 1094–5, 1100
control of 364
current solutions 1081–101
data analysis 1081–6, 1088, 1099
efficient sampling designs 1081–91
future developments 1101, 1114
medical devices 504–5
methodologic problems 1079–81
molecular pharmacoepidemiology 757, 760–1, 763, 764
pregnancy 537
randomized controlled trials 798
structures of cohorts 1081–2
study design 1079–101
unobserved patient characteristics 1095–101
vaccines 445–6
within-subject studies 1087–8
Congressional Office of Technology Assessment 639
conjoint analysis 878
consent 780–1, 783
consistency 48, 922
CONsolidated Standards Of Reporting Trials (CONSORT) guidelines 684
construct validity 597, 1071
constructive knowledge 144
consumer medication information 592
Consumer Product Safety Commission 707
consumer, definition 175–6
content validity 597
Contergan® 773
continuity corrections 917
continuous benefit–risk monitoring 135
contract law 140, 152–5
contract research 154, 159
control groups
metaanalyses 913
randomized controlled trials 801, 807
sample size 61–6, 68
see also case–control studies
control-to-case ratios 1084
controlled distribution products 345
Controlled Substance Act (CSA) 703
COPD *see* chronic obstructive pulmonary disease
copy number variants (CNV) 749
coronary heart disease 687
correlation 951
corticosteroids 296, 394, 1041
inhaled (ICS) 601, 1084
static asthmaticus 821, 1090
cost identification analysis 849
cost–benefit analysis (CBA) 847–8
cost-effectiveness analysis
beneficial drug effects 826–7
pharmacoeconomics 848–9, 850, 851, 858, 859
cost-effectiveness ratio 858, 859
cost–utility analysis 848
costs
clinical trials 71
direct medical costs 850
direct nonmedical costs 850
drug utilization 394
electronic health record databases 265
intangibles 850–1
pharmacoepidemiological studies 71
prescribing practices 412–13
productivity 845, 847, 850, 851
types of 850–1
see also pharmacoeconomics
cotrimoxazole 1035–40
Council for International Organizations of Medical Sciences (CIOMS) 173, 447, 592–3, 677
covariates 922
coverage 600, 641, 642
area 709–10
insurance 216,218, 766, 778, 844, 974
next-generation sequencing 766
period 214–15
prescription drug 838, 1116
risk minimization tool 597
Cox proportional-hazards modeling 507

Cox regression model 300, 534, 629
COX-2 *see* cyclo-oxygenase-2
COX-2 selective inhibitors 3311, 680, 1097
CPOE *see* computerized physician order entry
CPRD *see* Clinical Practice Research Datalink
CPT *see* Current Procedural Terminology
criterial methods 734
critical scored algorithms 735
CRNs *see* Clinical Research Networks (CRNs)
cross-sectional construct validity 597, 1071
cross-sectional studies 56
CRQ *see* Chronic Respiratory Questionnaire
crude pooling 687
CTCAE *see* Common Terminology Criteria for Adverse Events (CTCAE)
CT-P13 (infliximab-dyyb, Inflectra®) 569
CTSA *see* Clinical and Translational Science Awards
CUGCs *see* Clinical Utility Gene Cards (CUGCs)
cumulative metaanalysis 926–8, 928–9
cumulative revision rate over time 507
cumulative risk analysis 1101
cumulative sum of outcomes (CUSUM) 507
current procedural terminology (CPT) codes 213, 220, 979
CUSUM see cumulative sum of outcomes (CUSUM) methodology
CYP 3A4 inhibitors 31, 34, 37
CYP *see* cytochrome P450
cytochrome P450 enzymes (CYP)
  molecular pharmacoepidemiology 748, 750–1, 754–5
  pharmacokinetics 31, 34, 36, 40

d

Dabigatran 191–2
DAD *see* Discharge Abstract Database
Danish Medicines Agency (DKMA) 641
DAS-28 (disease activity score) 1070
data analysis
  confounding data 1081–6, 1088, 1099
  event monitoring 318–27
  large simple trial 801–2
  patient compliance/adherence 994, 1001
  vaccines 445–6
Data and Safety Monitoring Boards (DSMB) 447
data collection
  large simple trial 797, 801, 803
  medical devices 504
  metaanalyses 900, 910–12
  patient adherence/compliance 998–9
  validity of data 951
data dredging 682
Data Extraction and Longitudinal Trend Analysis (DELTA) system 508
data integration 136
data integrity 782
data linkages 508–9
  deterministic (direct) 508
  longitudinal data 508
  medical devices long-term follow-up 508–9
  probabilistic (indirect) 508
  vaccines 461–5
data mining 675–94
  biomedical computing methods 690–1
  causality and 681–3
  clinical problems 677–81
  currently available solutions 685–91
  exploratory and confirmatory analyses 684–5
  future 692–4
  in healthcare databases and distributed database networks 686–8
  medical literature 690
  medical devices 506
  methodologic problems 681–91
  replication and transparency in improving reliability 683–4
  rofecoxib, withdrawal of 680–1
  Sentinel Initiative 680–1, 683, 688, 692
  signal detection 676, 677–8
  signal evaluation 676
  signal refinement 676
  signal substantiation through proteins 679
  social media 688–90
  spontaneous reporting 187, 189, 190, 194–6
  spontaneous reports 685–6
  vaccines 464, 471, 472–3
data monitoring 802
data quality
  electronic health record databases 255–7, 259
  inpatient databases 294
  molecular pharmacoepidemiology 787
  vaccines 445–6, 450
Data Safety Monitoring Boards 783
data sources
  biologics 573–4
  comparative effectiveness research 652–5
  drug safety evaluations 102–3
  drug utilization 385–7
  metaanalyses 905, 908
  validity of data 950, 951
  see also data sources, choosing among
data sources, choosing among 357–70
  available approaches 357–60
  comparative characteristics 358–9, 360–5
  examples 369–70
  relative cost 360
  relative size of database 360, 363
  relative speed 360
  representativeness 360–3
  research question characteristics 365–9
Daubert v. Merrell Dow 147–8
DAWN *see* Drug Abuse Warning Network (DAWN)
days of therapy (DOT) 389
DDD *see* defined daily dose
DDNs *see* distributed data networks (DDNs)
Dear Healthcare Professional Letters
  risk management 591–2
  spontaneous reporting 172, 177, 195
Death Master File (DMF) 258
dechallenge 726, 732, 734
DEcIDE *see* Developing Evidence to Inform Decisions about Effectiveness (DEcIDE)

deep vein thromboses (DVT) 753
defined daily dose (DDD) 388–90, 392
degrees of safety 76
delayed risks 77
DELTA *see* Data Extraction and Longitudinal Trend Analysis
demographic data, validity of 962, 965, 966, 969
Department of Veterans' Affairs (VA)
  comparative effectiveness research 656
  Corporate Data Warehouse (CDW) 258
  databases 262, 264
  electronic health record database 257–9
  Healthcare System 806–7
  prescribing practices 413
  validity of data 979
depression 132, 143, 147, 536, 955
descriptive epidemiology 104
DESI *see* Drug Efficacy Study Implementation (DESI)
design defects 141–2, 143
Desirability of Outcome Ranking (DOOR) 888–9
deterministic linkage 217, 345, 508
developing countries 14, 15, 181, 195
Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) Network 12
Development Safety Update Report (DSUR) 447
device variation 513
dexamethasone 640
dexfenfluramine (Redux®) 141
diabetes
  children 536, 541
  data collection 344, 345
  encounter database 233, 234
  event monitoring 332
  gestational 538–9
  pregnant women 535, 537, 538–9
  prescribing practices 85, 86
  type 2 (NIDDM) 332, 344, 923, 931
  vaccine 472
diagnosis data 959–79
Diagnosis Related Group (DRG) 206–7, 462
diagnosis-linked databases 385
Diagnostic and Statistical Manual of Mental Disorders (DSM-V) 701
diethylstilbestrol (DES) 7, 17, 775
differential misclassification 953
differential opportunity for exposure 533
diphenhydramine 566
diphtheria, tetanus, acellular pertussis (DTaP) vaccine 441, 444, 457, 461
diphtheria–tetanus–pertussis (DTP) vaccine 444
direct medical costs 850
direct nonmedical costs 850
direct observation 1047, 1049
direct-to-consumer advertising (DTCA) 147, 151, 1022
Discharge Abstract Database (DAD) 226
discordance, patient/clinical/researcher 1069–70
discounting 77
DISCOVER study 345
disease-modifying antirheumatic drugs (DMARD) 1084
disease-oriented indicators 394
disproportionalities 184–5, 194
distributed data networks (DDNs)
  analysis 631
  challenges in 623–4
  clinical problems 621–3
  confounder summary scores 628–9
  currently available solutions 624–30
  data sources 631
  definition 617–18
  distributed regression 630
  encryption 630
  examples 618–21
  interoperability and coordination 632
  meta-analysis 629–30
  methodologic problems 623
  methodological advances 628–30
  postmarket surveillance 623
  rare exposures assessment 621
  rare outcomes assessment 621–2
  scope 631
  sustainability and efficiency 630–1
  transparency and reproducibility 631–2
  treatment effect heterogeneity assessment 622–3
distributed data systems 971–2
distribution 29
distribution-based methods 1071
DKMA *see* Danish Medicines Agency
DMARD *see* disease-modifying antirheumatic drugs
doctor/pharmacy shopping analyses 711
dose–response relationships
  clinical epidemiology 48–9
  risk factors 1053
DRG *see* Diagnosis Related Group
drug abuse 701–17
  analytic studies 703, 704, 705–9, 709–12
  clinical problems 703–13
  currently available solutions 713–16
  data sources 705–9
  deaths 707–8
  definitions 701–3
  diversion 708
  future 716–17
  methodologic problems 704–13
  quantitative descriptive assessment 703, 704, 705–9, 709–11,714–16
  signal detection 703, 705, 709, 713–14
  Type A/B reactions 703
  utilization data 709
Drug Abuse Warning Network (DAWN) 707, 714
drug approval processes 13–15
  other countries 14–15
  regulatory agencies 129
  US 13–14
drug class 75, 1053
drug concentration testing 1000
drug crises 7–13
drug development programs 127–30, 900, 915
drug discovery 765
drug–drug interactions (DDI) 1030–43
  clinical problems 1031
  current solutions 1032–3
  expected outcome 1030
  false warnings 1031
  future of 1043
  initiation order of object and precipitant 1040
  medication errors 1049
  methodologic problems 1031–2
  outcome assessment methods 1039

drug–drug interactions (DDI) (*cont'd*)
positive control pair to assess assay sensitivity 1039
research designs for studying health effects 1032–9
risk management 588, 589
screening to identify potential DDIs 1043
time course of 1040–3
drug efficacy 366
Drug Efficacy Study Implementation (DESI) 7, 11
Drug Enforcement Administration 714
Drug Epidemiology Unit 11
drug exposure 726–7
Drug-Induced Mortality database 708
Drug Information Number (DIN) (Canada) 226
drug interactions
future developments 1115
pharmacokinetics 37
risk management 588, 589
spontaneous reporting 173
drug labels *see* labeling
drug misuse 78
definition 701
Drug Safety and Effectiveness Network (DSEN), Canada 618, 934
Drug Safety and Risk Management Advisory Committee 12
Drug Safety Communications (DSCs) 169
drug safety evaluations
biopharmaceutical industry 101–12
data sources 102–3
postapproval safety studies 106–8
preapproval safety studies 103–6
special populations 108–12
study design 101–2
Drug Safety Research Trust 11
Drug Safety Research Unit (DSRU) 308, 309, 310, 311, 316, 318, 785
Drug Surveillance Research Unit 11
drug use results surveys (DURS) 100
drug utilization 375–99
antibiotics 376–7, 395
classification systems 389–92
clinical problems 378–80
currently available solutions 387–97
data sources 385–7
definitions 377–8
event monitoring 325–7, 333–4
future 397–400
historical background 375–7
intervention strategies 396–7
methodologic problems 380–7
pregnancy 526–7
prescribing practices 412–13
quality indicators 392–6
study designs 381–2
units of measurement 388–9
drug utilization review (DUR) programs 377, 381, 397, 413, 416, 423, 1112, 1114
DSEN *see* Drug Safety and Effectiveness Network (DSEN), Canada
DSMB *see* Data and Safety Monitoring Boards
DSRU *see* Drug Safety Research Unit (DSRU)
DSUR *see* Development Safety Update Report
DTCA *see* direct-to-consumer advertising
DTP *see* diphtheria–tetanus–pertussis
DUR *see* drug utilization review
duration–response relationships 48
Dutch National Pharmacovigilance Center Lareb 176
DVT *see* deep vein thromboses
dwell time 215

e

EBGM *see* empirical Bayes geometric mean (EBGM)
ECLAMC (Latin-American Collaborative Study of Congenital Malformations) network 545
ecologic time-series assessments 709
ecological studies *see* analyses of secular trends
economic evaluations *see* pharmacoeconomics
EDM *see* electronic drug monitors
educational materials 419–20
effect modifications
children 538
metaanalysis 922
molecular pharmacoepidemiology 757–8
pregnancy 532
Effective Practice and Organisation of Care (EPOC) 412, 414
effectiveness
beneficial drug effects 814, 818, 822, 825
definition 641
future developments 1114
medical devices 502–3
metaanalyses 932
prescribing practices 411, 412, 413, 414
risk evaluation studies 1013, 1015
risk management 587, 589
see also comparative effectiveness research
effectiveness of medication, definition 88
effectiveness research *see* comparative effectiveness research
efficacy of medication, definition 88
efficacy studies
beneficial drug effects 814
metaanalyses 929, 930
risk management 589
EHRs *see* electronic health records databases
elderly populations
inpatient databases 298–9
medication errors 1051
pharmacodynamics 33
pharmacokinetics 32–4
electrocardiographic QT interval 1032
electronic databases 102–3, 205–10
administrative databases 206–7
claims data 206–7
future 209–10
limitations 230
sample size 70
strengths 207–8, 230
vaccines 461–5
weaknesses 208–9
see also electronic health record databases
electronic drug monitors (EDM) 995, 999
Electronic Frontier Foundation (EFF) 787
electronic health record databases 207, 241–66, 806
applications 265
drug abuse 707
Europe and the United Kingdom 252–7
examples of studies 260–1
future 265–6

limitation 263–5
medication errors 1047–8
overview 242–52
pharmacovigilance and 678
strengths 259–63
US: Department of Veteran Affairs Healthcare 257–9
validity of data 948
variables in 246–51
electronic medical record (EMR) databases 115
elements to assure safe use (ETASU) 605, 606, 1020–1
Eli Lilly 345
elimination 32
empirical Bayes geometric mean (EBGM) 685
EMR *see* electronic medical records
EMRS *see* European Medicines Regulatory Network
enabling strategies 420
ENCePP *see* European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
encounter databases 211–35
ability to validate outcomes 233–4
access 217–18
applications 232–4
Asia 228
attributes 214–15
Canada 225–6
characteristics 219
data domains 212–13
database population 214
database size 233
description 212–14
European 228
federated databases 213
future 235
integrated databases 213
limitations 231–2
linkage to nonencounter data 216–17
nonstandard encounter data 234
Nordic 227–8
population and coverage period 214–15
selected databases 218–29
services covered and data completeness 215–16
size of database 214
strengths 229–31
target population 233
US 218–25
US government 222–5
US private insurance databases 220–2
validation of 213–14
endpoints 343, 346, 380
enhanced safety surveillance (ESS) 344, 345
enhanced surveillance 506–7
Enhancing the QUAlity and Transparency Of health Research" (EQUATOR) network 684
enrollment logs 349
EPAR *see* European Public Assessment Report
EPhMRA *see* European Pharmaceutical Market Research Association
Epic Database Research Company Ltd 253
epidemiology, definition 5
see also clinical epidemiology
EPOC *see* Effective Practice and Organisation of Care
EQUATOR network 684
errors
clinical epidemiology 44, 45, 46–7
coding 626
see also measurement error; medication errors; Type I/II errors
etanercept 569
ETASU *see* Elements to Assure Safe Use
ethics 772–87
clinical problems 773–80
conflicts of interest 782–3, 784–5
current solutions 783–5
data integrity 782
East Asian 778–80
European 776–8
future 785–7
informed consent 780–1, 783
methodologic problems 780 3
North American, 20th century 773–6
pharmacovigilance practices 783–4
surveillance 781–2
ethnicity *see* race/ethnicity
EudraVigilance database 169, 183, 193, 584, 777
EU-PACT UK trial 755–6
EUROCARE 350
EUROCOURSE 350
EUROmediCAT 545
European electronic health record databases 252–7
European Cancer Observatory (ECO) 350
European Commission 777
European Drug Utilization Research Group (EuroDURG) 390, 399, 776
European Medicines Agency (EMA) 15, 40
adaptive pathways 85
assessment reports 867
benefit–risk assessment 867
Committee for Medical Products for Human Use 109
drug labeling 751
drug safety evaluations 103, 106
enhanced safety surveillance 344
founding of 776–7
Good Pharmacovigilance Practices (GVP) 778, 782, 783, 784
influenza vaccines 587
M-PEM studies308–11, 317
Pharmacovigilance and Risk Assessment Committee (PRAC) 100, 778
PRIME initiative 128
randomized controlled trials 794
relocation of 786
risk management 99, 587
vaccine safety 440
European Medicines Regulatory Network (EMRN) 777
European Network of Cancer Registries (ENCR) 350
European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) 92, 103, 134, 136, 337, 347, 785, 904, 909, 934
Checklist for Study Protocols 778
Guide on Methodological Standards in Pharmacoepidemiology 778, 784
European Pharmaceutical Market Research Association (EPhMRA) 390
European Public Assessment Report (EPAR) 594, 1014–15

European Union
Clinical Trials Regulation 784
Exploring and Understanding Adverse Drug Reactions (EU-ADR) 621, 630. 678, 680
Innovative Medicines Initiative (IMI) "WEB-RADR" project 689–90
pharmacoepidemiologic postauthorization safety studies (PASS) 100, 337, 783–4
postauthorization study (PAS) register, hosted by the ENCePP 784
euthanasia 78
event, definition 312
event adjudication 926
event monitoring 307–38
analysis vs event counts (incidence) 319
applications 330–7
automatic signal generation 333
before and after studies 332
bias 329–30
children 334
comparison of event rates and risks 331
confounding 330
data analysis 318–27
data collection 311–12
data processing 316
definitive answers 328
design and source data 308–10
drug utilization 325–7, 333–4
electronic data capture 337
ethics and confidentiality 310–11
exposure data 327
future 337–8
incidence densities (IDs) 319–25
modeling 332–3
multicountry network studies 337–8
outcome data 327–8
participation in research 328–9
pharmacovigilance risk management plans 335–7
pregnancy 325
qualitative evaluation of adverse events 318
quantifying AD reporting 334–5
quantitative analysis of events 318–19
questionnaire design 312
reference (contextual) cohorts 328
representativeness and size 327
sample size and duration 316–18
signal strengthening 328, 330–1
simple stratification 331–2
single-group cohort design 329
standardization 333
statistical power and sample size 330
stopping the drug 325
supplemental information 312
variable direction of investigation 327
see also modified Prescription Event Monitoring (M-PEM); specialist cohort event monitoring (SCEM)
evidence-based practice
comparative effectiveness research 662–3
medical devices 501
metaanalyses 901
prescribing practices 419
evidence-based research *see* comparative effectiveness research
excess risk 54, 76–7
execution 992, 995
experimental studies *see* randomized controlled trials
expert witnesses 140, 147
expertise 147–8
exploratory data analysis (EDA) 683
exposure data 363–4
beneficial drug effects 820
event monitoring 327
medical devices 499, 503–4
new drug 368
vaccines 444, 465–6
exposure frequency 544
exposure misclassification
confounding data 1101
vaccines 444
validity of data 973

f

face validity 957
factor V Leiden mutations 753
FAERS *see under* Food and Drug Administration
fail-first policies 572
fail-safe *N* approach 918–19
failure to warn claims 142, 143–6
causation of damages 145–6
improper warning of drug risk 144–5
knowledge of drug risk by manufacturer 143–4
false discovery rate (FDR) 764
false positives/negatives *see* Type I/II errors
FDA *see* Food and Drug Administration
Federal Commission for Protection against Health Risks (COFEPRIS) (Mexico) 100
Federal Institute for Drugs and Medical Devices (BfArM) 182
Federal Rules of Evidence 148
feedback systems 422–3
fenofibrate 1036
FEP *see* Federal Employee Program
FFS *see* fee-for-service
finasteride 319, 320–1
Fingolimod 190–1
fluoxetine 7, 143–4
follow-up
health maintenance organizations 797
large simple trials 801
medical devices 508–9
follow-up-time cohorts 1082
Food and Drug Administration (FDA)
Adverse Event Reporting System (FAERS) 169, 705, 713
beneficial drug effects 815, 816, 820
benefit–risk assessment 867
causality 727–9
drug abuse 701
drug development program 128, 129, 131
drug marketing 6–7
drug safety 35, 106, 109
elderly populations 34
ethics 773
future developments 1117
historical development 6–7, 10–13
legislative instruments 143–5, 148–50
medical devices 506, 511
metaanalyses 904, 905, 920, 923, 924–5, 928, 930–1
model-informed drug development (MIDD) 40
pharmacoeconomics 838
pharmacoepidemiology applications 15–16
pharmacokinetics 35
postmarketing research 72
Pregnancy and Lactation Labeling Rule (PLLR) 529
Prescription Drug User Fee 785

product liability legal theory 142
regulatory factors 72–3
risk evaluation studies 1010, 1018–21
risk management 581, 604–5, 609, 611
risk mitigation 112–13
spontaneous reporting 176–7, 194
Spontaneous Reporting System 73
vaccines 440, 448–9, 460, 463
Voice of the Patient reports 871
see also Sentinel Initiative
Food and Drug Administration Amendment Act (FDAAA) (2007) 99, 776
Food and Drug Administration Center for Drug Evaluation and Research (FDA CDER) 349
Food and Drug Administration Modernization Act (FDAMA) (1997) 9, 11, 511, 530
Food and Drug Administration Reauthorization Act (2017) (FDARA) 10
Food and Drug Administration Safety and Innovation Act (2012) (FDASIA) 10
Food, Drug, and Cosmetic Act (1938) 6
formulation effects 75, 178
French Medicines Agency 182
funding 660–1, 1117
funnel plots 507–8, 918–20
fuzzy-trace theory 1023

g

G theory *see* generalizability
G6PD *see* glucose-6-phosphate dehydrogenase deficiency
gag clauses 153
gap rule 997
Gardasil® 681
gastrointestinal (GI) bleeding
confounding data 1097
validity of data 966, 967–8, 975–8
gastrointestinal (GI) pharmacokinetics 35
Gastrointestinal Toxicity Survey 967
GBS *see* Guillain–Barré syndrome
Gemscript 254
gender effects
cultural factors 379
validity of data 962, 966
gene discovery approaches 761–2
gene–drug interactions 750–1
general causation 145–6
General Practice Research Database® (GPRD®) 252
advantage of 387
comparative effectiveness research 656
data utilization 387
study cohort 230
validity of data 976
generalizability
comparative effectiveness research 653
medical devices 505, 512
pharmacoeconomics 846
randomized controlled trials 797
Generic Drug User Fee Amendments (GDUFA) (2012) 10
generic drugs 825–6
genetic variability 747–8, 753
genome-wide association studies (GWAS) 753, 761–2, 764
geriatrics *see* elderly populations
GFR *see* glomerular filtration
GIFT trial 756
Gilead 844
glaucoma 195
GlaxoSmithKline 345
Global Harmonization Task Force (GHTF) 497
global introspection 728, 733
Global Research in Paediatrics (GRIP) 621
Network of Excellence 544
glomerular filtration rate (GFR) 32, 34
glucose-6-phosphate dehydrogenase (G6PD) deficiency 747
glucuronosyltransferase (UGT) isoforms 34
glutethimide 534
glyburide 1035
gold standard 949, 950, 951
golimumab 569
governance 116, 623, 624, 630, 632, 654–6
clinical 641
see also ethics
government claims databases *see* Medicare; Medicaid
GPRD® *see* General Practice Research Database®
GRACE (Good Research for Comparative Effectiveness) Principles 347
Green Forms 311, 312, 313
Group Health Cooperative 776, 979
guideline materials 419–20
Guidelines for Good Pharmacoepidemiologic Practice 347
Guillain–Barré syndrome (GBS) 438, 439, 444–5, 451, 453, 454, 455
GUSTO-I/GUSTO-IIb 825
gynecomastia 319, 320–1

h

Harvard Pilgrim Healthcare Institute 933
HAS *see* Haute Authorité de Santé
Haute Authorité de Santé (HAS) 640, 641
Hawthorne effect 994
HCPS 571
HCSRN *see* Health Care Systems Research Network (HCSRN)
HDL *see* high density lipoprotein
Health Care Systems Research Network (HCSRN) 618–19, 624
health insurance 838–9, 839–40
see also commercial insurance databases
Health Insurance Portability and Accountability Act (HIPAA) 220, 291
compliant databases 294
health information 787
Health Maintenance Organization Research Network (HMORN) 618–20
health maintenance organizations (HMOs)
accuracy 964
bioethics 776
data 207, 365, 368
epidemiology 103
follow-up data 797
pharmacoeconomics 845
postapproval surveillance 461
postmarketing study 845
size 360
Health Measures 1073
health plans *see* health maintenance organizations
health professional/consumer medication information 1014–21

health technology assessment *see* comparative effectiveness research
Healthcare Common Procedure Coding System (HCPCS) 220, 569, 570
  healthcare databases 680–1
healthcare databases 680–1
healthcare providers (HCPs) 346
healthcare systems 83–93
  assessment of quality and outcomes of drug use in populations 88–9
  drug approval process 84–5
  economic assessment of medication-related issues 91
  evaluation of use of drugs 86–8
  future 92–3
  interventional pharmacoepidemiology 90–1
  policy analysis 89–90
  prescribing practices 85–6, 413
  risk management 591–2
  spontaneous reporting 175–7, 181, 188
  see also health maintenance organizations
HealthCore® Integrated Research Database (HIRDSM) 974, 978, 979
Henoch–Schönlein purpura (HSP) 296–8
Heparin 178, 296, 911
hepatic function 34–5
hepatitis
  cohort studies 63
hepatitis B
  biologics 564
  vaccine safety 437, 444, 453, 456, 460, 466
hepatitis C
  pharmacoeconomics 844–5
  VA database 258
hepatotoxicity 1011, 1016
  case–control studies 65
  cohort studies 63
hepatic function 34–5
HER2 oncogene 753
herbal medications 376, 980, 1114
heterogeneity 913–15, 915–16, 932
high-density lipoprotein (HDL) cholesterol 1093–4
HIPAA *see* Health Insurance Portability and Accountability Act
HIRA 688
historical background 6–13
  early legislation 6–7
  drug crises 7–13
  intellectual development 10–13
  legislation and regulations 6–7, 9–10
  legislation actions 9–10
  regulatory actions 7–9
HIV/AIDS 716, 793
  patient compliance/adherence 991, 993, 995, 997, 1000, 1002
  pregnancy exposure registries 109
HMO *see* health maintenance organizations
homogeneity of data 921–2
hormone replacement therapy (HRT) 12–13, 55, 822, 869, 903
hospitalization data
  confounding data 1083, 1090
  patient compliance/adherence 991
  validity 963–6, 973–4
  see also in-hospital databases
Howard Hughes Medical Institute 837
HRT *see* hormone replacement therapy
HSP *see* Henoch–Schönlein purpura
human capital development 661–3
Humana 360
hydrochlorothiazide (HCTZ) 147
hyperbolic concentration effects 38
hypercholesterolemia 85, 87, 173
hypergeometric distribution 156
hypersensitivity reactions
  allergic 369–70
  α-chymotrypsin 195
  biologics 566
  definition 5
  heparin 178
  NSAIDs 818
  spontaneous reporting 178
  vaccines 453
hypothesis generating studies 75–6, 365
hypothesis strengthening 365–6, 452, 459, 1116
hypothesis testing 75, 366
  event monitoring 330, 331–2
  large simple trials 799
hypothesis-generating studies 75–6
hysteresis 38–9

i

I^2 statistic 915–16
IAAAS *see* International Agranulocytosis and Aplastic Anemia Study
IAHPR *see* International Academy of Health Preference Research
IARC *see* International Agency for Research on Cancer
ibuprofen 66, 67, 799
  large simple trials 802–4
ICH *see* International Conference on Harmonization
ICMJE *see* International Committee for Medical Journal Editors
IDIS *see* Iowa Drug Information Service
IIS *see* Immunization Information Systems
IL-17 therapy 567
illicit drugs 107, 701, 703, 706, 712, 717
imatinib 753, 756, 852–7
IMDRF *see* International Medical Device Regulators Forum
IMI PROTECT 185
IMI *see* Innovative Medicines Initiative
immediate risks 77
Immunization Information Systems (IIS) 466
Immunization Monitoring Program–Active (IMPACT) 449
immunizations *see* vaccines
IMPACT *see* Immunization Monitoring Program-Active
implantable medical devices *see* medical devices
implementation 991
in-hospital databases 215, 364
  see also inpatient databases
in utero exposure
  historical development 7
  pharmacoepidemiology applications 17
  see also birth defects; pregnancy
inactivated polio vaccine (IPV) 439, 440,457, 461
inception cohort design 1040
incidence
  disease 366
  event monitoring 319
  medication errors 1050
  sample size 62–3, 65, 66–8
incidence densities (IDs) 319–25
incidence rate ratio (IRR) 333, 467–8, 1035
incident reports 1047, 1048
incident user design 1080, 1095
incomplete data
  validity 960, 966, 969

incremental cost effectiveness ratio 849
independent medical assessments 178–9
indication-specific questions 962
indirect comparisons 921–2, 930
individualization of therapy 4
indomethacin 7
intelligent nonadherence 86–7
infections 4, 17
see also antibiotics
infliximab 564, 568, 569, 570, 571
information bias 329, 1090
information component (IC) 685
informed consent 780–1, 783
ethics 775
initiation 991
Inman, Professor William 308
Innovation in Medical Evidence Development and Surveillance (IMEDS) 631, 933, 934
Innovative Medicines Initiative (IMI) 116, 135, 180, 343, 464, 599, 620, 886
PREFER project 886
inpatient databases 290–302
adverse event evaluation and therapeutic monitoring 298–300
applications 295–302
clinical practices variation 296–8
common data elements 292
comparative effectiveness 300–2
comparison 293
data accessibility 294–5
data quality 294
description 291–4
disease status 295
exposure status 295
future 302
outcome status 295
strengths 294–5
temporal trends 296
weaknesses 295
inpatient diagnoses
medication errors 1053–4
validity of data 973–4
inpatient drug use 364, 367
InPreSS *see* International Pregnancy Safety Study
INRUD *see* International Network for Rational Use of Drugs
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen *see* IQWIG
Institute for Clinical and Economic Review (ICER) 845
Institute for Quality and Efficiency in Health Care *see* IQWIG
Institute of Medicine (IoM) 9–10
comparative effectiveness research 640, 650, 652, 658–9, 660
medication errors 1046
risk management 99
vaccines 438,452, 460, 461, 463, 471
institutional review boards (IRB)
ethics 780, 783
instrumental variables 820, 1096–9
insulin 7
insurance claims databases 115
insurance *see* commercial insurance databases
intangible costs 850–1
integrated healthcare delivery system databases 222
Integrated Primary Care Information (IPCI) 254, 387
intellectual property law 140, 155–7
in Europe 157–8
intelligent nonadherence 86
intensive care units (ICU) 1050, 1055
intensive monitoring 100
intent-to-treat (ITT) analysis 793
Interdisciplinary Processing of Clinical Information 254
interim analyses 802
internal validity 414–16
International Academy of Health Preference Research (IAHPR) 886
International Agency for Research on Cancer (IARC) 350
International Agranulocytosis and Aplastic Anemia Study (IAAAS) 1084
International Classification of Diseases (ICD) 707, 1039
vaccines 466
validity of data 974, 979
International Classification of Diseases, Clinical Modification (ICD-CM) system 220
International Committee for Medical Journal Editors (ICMJE) 154
International Conference on Harmonization (ICH) 170, 177, 530
benefit–risk assessment 867
Clinical Overview 886
International Council for Harmonisation (ICH) 870
International Epidemiological Association
Epidemiology Practice (GEP) guidelines 795
International Medical Device Regulators Forum (IMDRF) 497
International Network for Rational Use of Drugs (INRUD) 376, 396
international postmarketing safety databases 182–3
International Pregnancy Safety Study (InPreSS) 543
International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 117, 657, 684, 886
International Society for Pharmacoepidemiology (ISPE) 11, 347, 399, 1111, 1112
Guidelines for Good Pharmacoepidemiology Practice (GPP) 783, 784
International Warfarin Pharmacogenetics Consortium (IWPC) 764
Internet
data collection/entry 348, 349, 350
reports 174
risk evaluation 1022
interrater reliability 906
interrupted time-series designs 414
interventional studies 782
interventional pharmacoepidemiology 90–1
intraclass correlation coefficients 952
inverse-variance methods 913
investigational new drugs (IND) 6, 730
IoM *see* Institute of Medicine
Iowa Drug Information Service (IDIS) 390
IPCI *see* Integrated Primary Care Information
ipilimumab 566
iPLEDGE® 603–4
IPPHS 784
IPV *see* inactivated polio vaccine
IQVIA 253, 256
Disease Analyzer databases (DA) 253
National Sales Perspective database 386
IQWiG 640, 641, 845

IRB *see* institutional review board
IRR *see* incidence rate ratio
isotretinoin 7, 76, 78, 602–4, 681
ISPE *see* International Society for Pharmacoepidemiology
ISPOR *see* International Society for Pharmacoeconomics and Outcomes Research
item response theory (IRT) 1062, 1071–2
IVH *see* intraventricular hemorrhage
IWPC *see* International Warfarin Pharmacogenetics Consortium
ixazomib 640

j

Japan
  Medical Information for Risk Assessment Initiative (MIHARI) 100, 779
  Pharmaceuticals and Medical Devices Agency (PMDA) 100, 128, 779
  SAKIGAKE review program 128
Joinpoint regression 709
Joint Commission on Accreditation of Health Care Organizations (JCAHO) 735, 1114
Joint Commission on Prescription Drug Use 11, 775–6
Joint National Committee on Prevention, Detection, and Treatment of High Blood Pressure (JNC) 86
juvenile idiopathic arthritis (JIA) 101, 526, 536

k

Kaiser Permanente (KP) 221, 363, 369, 776, 824
Kaiser Permanente National Joint Replacement Registry (NJRR)503
Kaplan–Meier estimates 887
kappa statistic 952
Kefauver–Harris Amendments 6, 10–11, 529, 815, 1010
ketorolac 74, 75
kidney *see* renal function
known risks 583–5
Korea Food and Drug Administration (KFDA)
  pharmacovigilance research network (PVNet) 778
KP *see* Kaiser Permanente

l

LABA *see* long-acting beta agonists
labeling
  beneficial drug effects 815
  causation 730
  European Medicines Agency (EMA) 751
  legislative instruments 148–50, 151–2, 159
  pediatric medications 527
  postmarketing surveillance 73
  risk management 591–3, 601–2, 604, 605
  safety-related label changes 8
  spontaneous reporting 170
laboratory data
  baseline 726
  clinical assessments 341
  electronic health record databases 255, 258
  encounter databases 227
  inpatient databases 291
lactating women
  ethics and regulations 528–9
large simple trial (LST) 793, 794–5
  appropriateness of 797–8
  cooperative population 799
  data analysis 801–2
  data collection 801
  data monitoring/interim analyses 802
  design 796
  electronic data capture in 806–7
  feasibility 798–9
  follow-up 801
  future 808–9
  logistics of 800–1
  masking 795–7
  outcomes 799
  power and confounding 798
  primary analysis 801
  research question 798
  simple hypothesis 799
  simple treatment 799
  subgroup analyses 801–2
  uncertainty 798
LASSO (least absolute shrinkage and selection operator) 505
lasso shrinkage regression 686
law *see* legal factors; legislative instruments
LD *see* linkage disequilibrium
LDL *see* low density lipoprotein
learned intermediary defense 146–7, 151–2
Learning Healthcare System 91
legal benefits 74–5
legal factors
  causation 730–1
  future developments 1120
  pharmacoepidemiological studies 74–5, 76
  prescribing practices 417–18
legislative instruments 140–59
  contract law 140, 152–5
  expert witnesses 140, 147
  expertise and *Daubert* 147–8
  failure to warn claims 142, 143–6
  intellectual property law 140, 155–8
  learned intermediary defense 146–7
  patents 140, 155–8
  product liability lawsuits 140–8
  tort law 140–52
lenalidomide 640
lidocaine 824–5
lifestyle factors 965, 969
life-threatening side-effects 586
linear concentration effects 38
linkage disequilibrium (LD) 752, 755, 763
literature reviews 897, 902, 903–5
liver toxicity *see* hepatotoxicity
long-acting beta agonists (LABA) 601
longitudinal data 302
  administrative databases 512
  Bayesian approach 888
  confounding data 1078, 1079, 1095
  data linkages for 508
  group-based trajectory modeling 1103
  inpatient database 292
  validity 970
Longitudinal Evaluation of Observational Profiles of Adverse events Related to Drugs (LEOPARD) 688
longitudinal observational databases (LODs) 115
long-term care 1054
long-term safety 502–3
loperamide 713–14
loss to follow-up 365
low density lipoprotein (LDL) cholesterol 1093–4

low incidence outcome 366–7
low molecular weight heparin 296, 825, 911
low prevalence exposure 367
LSD *see* lysosomal storage disorders
LST *see* large simple trials
lumiracoxib 312, 314–15
lysosomal storage disorders (LSDs) 344

m

MAA *see* Marketing Authorization Applications
MAH *see* marketing authorization holder
major adverse cardiac events (MACEs) 929
Managed Care Organizations (MCO) 462–3, 470
mandated postmarket studies 511
Mantel–Haenszel procedure 913, 917, 925, 927, 929, 1088, 1089
Manufacturer and User Facility Device Experience (MAUDE) 506
marketing 73–4
Marketing Authorization Applications (MAA) 99
marketing authorization holder (MAH) 129, 182, 195, 318, 777
marketing benefits 73–4
MarketScan® 222, 230
Markov chain Monte Carlo methods 854, 930
masking 795–7
Master Beneficiary Summary Files (MBSFs) 224
MAUDE *see* Manufacturer and User Facility Device Experience
MAX data 223–4
maximized sequential probability ratio test (MaxSPRT) 463
MaxSPRT *see* maximized sequential probability ratio test (MaxSPRT)
MBSFs *see* Master Beneficiary Summary Files (MBSFs)
MCO *see* Managed Care Organizations (MCO)
MDEpiNet *see* Medical Devices Epidemiology Network
MDIC *see* Medical Device Innovation Consortium (MDIC)
"me too" drugs 75
measles–mumps–rubella–varicella (MMRV) vaccine 455, 461, 463
measurement error 949, 952–5
  indices of 949
MEB *see* Medicines Evaluation Board
mechanism of action (MOA) 561, 567
MedDRA *see* Medical Dictionary for Regulatory Activities
MED-ECHO 226
media reporting 425
Mediator® (benfluorex) 777
Medicaid 11, 222–4, 364, 365, 368, 542, 544, 776
  beneficial drug effects 826
  drug utilization 387
  validity of data 955
Medicaid Analytic eXtract (MAX) data files 215, 223
Medicaid Statistical Information System (MSIS) 223
  Transformed-MSIS (T-MSIS) 223
Medical Device Innovation Consortium (MDIC) 886
medical devices 496–517
  active surveillance782
  administrative claims data 511–13
  automated surveillance 508
  benefit–risk profiles 496, 498
  characteristics compared with drugs 498–501
  classification 498–9
  clinical practice and utilization 501
  clinical problems 501–3
  comparability of study groups 504–5
  comparative effectiveness and safety estimation 514–15
  comparative studies 504
  currently available solutions 506–15
  data linkages for long-term follow-up 508–9
  definitions 496–8
  enhanced surveillance 506–7
  future 515–16
  individual patient exposure 503
  international infrastructure 516
  long-term outcomes 505–6
  long-term safety and effectiveness 502–3
  mandated postmarket studies 511
  methodologic problems 503–6
  national population exposure 503–4
  outcome evaluation methods 513
  passive surveillance 506
  performance evaluation 501–2
  registries 509–11
  sample size 505–6
  signal detection/outlier identification 507–8
  simultaneous combination of available evidence 514–15
  sources of variation 513
  treatment assignment mechanisms 513–15
Medical Devices Epidemiology Network (MDEpiNet) Initiative 504, 515–16
  National Medical Device Registry Task Force Report 509
Medical Dictionary for Regulatory Activities (MedDRA) 183, 194, 316, 449, 459, 690, 691
Medical Information for Risk Assessment Initiative (MIHARI) project 100, 779
medical literature 690
Medical Product Safety Network (MedSun) 506
medical records
  mandated postmarket studies 511
  medication errors 1047–8
  validity of data 959–61, 965–6, 969,970, 973–4, 979
  see also electronic health record databases
Medicare 222–5, 367, 776
  confounding data 1099
  database relative size 360
  future developments 1116
  pharmacoeconomics 858
  representativeness 363
  validity of data 976
Medicare Advantage 224
Medicare Prescription Drug, Improvement, and Modernization Act (2003) 12
medication adherence, pregnancy 526–7
medication diaries 998–9
medication errors 1046–56
  clinical problems 1046–9
  current solutions 1054–6
  definition and classification 1046–7
  detection 1047–9, 1051–2
  drug class 1053

medication errors (*cont'd*)
future developments 1056
health-related outcome impacts 1052
incidence rates 1050
methodologic problems 1049–54
prescribing practices 416, 424, 428
prevention strategies 1054–5
risk factors 1052–3
risk management 582
settings 1050–1, 1053–4
special populations 1051
spontaneous reporting 178, 193–4
Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP) 543
medication possession ratio (MPR) 997
Medicine Monitoring Unit (MEMO) database 360
Medicines Act 307
Medicines and Healthcare Products Regulatory Agency (MHRA) 15, 183, 252
Medicines Evaluation Board (MEB) 182
Mediplus® 253
MEDLINE database 690
MedSun *see* Medical Product Safety Network
MedWatch 8, 10, 169, 171, 175, 177, 193, 506, 1020
membership status 213, 215
MEMO *see* Medicine Monitoring Unit database
menarche 966
meningitis 440
aseptic 440, 441,453 458
meningococcal conjugate vaccine 445, 455, 456,460, 461, 463
menopause 208, 209, 822, 966
Menstrual and Reproductive Health Study 966
Mental Health Research Network 618
MEPREP *see* Medication Exposure in Pregnancy Risk Evaluation Program
metaanalyses 897–934
adjusting for covariates922
adverse events/adverse drug reactions 898, 899, 900, 903, 922–3, 924, 926, 929, 930
assumptions 921–2
biases in data abstraction 905–6, 910
biases in original studies 902
case studies of applications 921–31
choice of statistical test 912–15
clinical problems 899–901
combining studies 902–3, 915–16
comparative effectiveness research 649–50
conclusions and recommendations 918
cumulative 926–8, 928–9
current solutions 906–31
data collection 900, 910–12
definitions and rationale 897–9
differential effects among patient subgroups 924–6
drug approval process 84
future developments 931–4
inclusion/exclusion criteria 908–10, 929
indirect and simultaneous comparisons of treatments 921–2, 930
literature reviews 897, 902, 903–5
medical devices 505
methodologic problems 902–6, 920, 921–2
molecular pharmacoepidemiology 763, 764
new indications for existing therapies 924
publication bias 903–5, 918–20, 930, 931
purpose definition 907–8
rare events 916–17
regulatory factors 930–1
signal detection 928–9
statistical analyses 912–15
steps involved 907–20
time and resources savings 926–8
unpublished studies 918–20
metabolism 30–1
metaepidemiologic studies 764
metastatic breast cancer 344, 348, 753
metformin 1039
methicillin-resistant *Staphylococcus aureus* (MRSA) 291
methotrexate 567
methylphenidate 193, 233, 649–50
metiamide 75
MHRA *see* Medicines and Healthcare Products Regulatory Agency
MHRN *see* Mental Health Research Network
Michaelis–Menten equation 30
MID *see* minimally important difference
Mini-Sentinel 365, 620, 688
minimal clinically important difference (MCID) 1063, 1071
minimal important difference (MID) 1071
Minimum Data Set (MDS) 234
misclassification errors 952–3
see also exposure misclassification
MiSeqDx 762
missing doses 998
MMR *see* mumps–measles–rubella
MMRV *see* measles–mumps–rubella–varicella (MMRV)
mobile health (mHealth) technology 1066
modified Prescription Event Monitoring (M-PEM) 308–12, 317
molecular pharmacoepidemiology 746–66
applications 754–7
candidate gene approaches 761–2
case-only studies 762–3
causal pathway of disease 753
clinical problems 750–4
confounding data 757, 760–1, 764
current solutions 761–5
definitions and concepts 748–9
effect modifications 757–8
future developments 765–6
gene discovery approaches 761–2
gene–drug interactions 750–4
genetic variability 747–8, 748–9
historical development 747–8
methodologic problems 757–61
pharmacodynamics 752–3
pharmacogenetics and pharmacogenomics 749
pharmacokinetics 750–2
population admixture 760–1, 765
progression and clinical application 754–7
Type I/II errors 759–60, 763–5
Monitoring System for Adverse Events Following Immunizations (MSAEFI) 457
Monitoring the Future (MTF) 705–6, 714
moral hazard 839–41

Morisky Medication Adherence Questionnaire (MMAS-8) 995
morphine 710
motor vehicle crashes 1090
M-PEM *see* modified Prescription Event Monitoring (M-PEM)
MPR *see* medication possession ratio (MPR)
MRSA *see* methicillin-resistant *Staphylococcus aureus*
MSAEFI *see* Monitoring System for Adverse Events Following Immunizations
MTF *see* Monitoring the Future (MTF)
multicriteria decision analysis (MCDA) 870, 884, 885
multimedia campaigns 425
multiple sclerosis (MS) 114, 190–1, 564, 609, 884
multi-time case–control studies 1084–5
multivariable modeling 331
multivariate analysis 801, 1092–5, 1998 1099
mumps–measles–rubella (MMR) vaccine 437, 440, 457, 463
myocardial infarction (MI)
  beneficial drug effects 824–5
  metaanalyses 898, 901,909, 923
  validity of data 953, 964, 966, 967–8, 975–9
Mysimba 611–12

n

N-acetyl-*p*-benzoquinone imine 31
NAEPP *see* National Asthma Education and Prevention Program
naloxone 818
naltrexone 611–12
Naranjo scoring method 735
natalizumab 114–15, 564, 677, 679, 884–5
National Academy of Medicine *see* Institute of Medicine
National Addictions Vigilance Intervention and Prevention Program (NAVIPPRO®) 715
National Asthma Education and Prevention Program (NAEPP) 601
National Birth Defects Prevention Study 545
National Breast Implant Registry (NBIR) 503
National Cancer Institute (NCI) 347
National Cardiovascular Data Registry 345, 503
National Center for Health Care Technology 639
National Childhood Vaccine Injury Act 448, 465
National Comprehensive Cancer Centre (NCCN) Task Force 756
National Coordinating Council on Medication Error Reporting and Prevention (NCCMERP) 584
National Death Index (NDI) 349, 708,797, 801
National Disease and Therapeutic Index (NDTI) 387
National Drug Codes (NDC) 570, 626
  beneficial drug effects 826
  medical devices 503
  medication data 220
  National Electronic Injury Surveillance System: Cooperative Adverse Drug Events Surveillance System (NEISS-CADES) surveillance network 707
National Evaluation System for health Technology (NEST) Coordinating Center 504, 511
National Health Service (NHS) 845
National Health Service Business Services Authority (NHSBSA) 309, 311
National Health Service Organization Data Services (ODS) 309
National Hospital Ambulatory Medical Care Survey (NHAMCS) 601
National Hospital Discharge Survey (NHDS) 601
National Information Governance Board 310
National Institute for Health and Clinical Excellence (NICE) 12, 13, 640–1
National Institute of Health Research (NIHR) 252, 310
National Institutes of Health (NIH) 837, 1117, 118, 119
  NIH Toolbox® 1073
  Precision Medicine Initiative 843
National Medical Device Postmarket Surveillance System Planning Board 511
National Medical Device Registries Task Force 511
National Multiple Cause of Death Files 601
National Patient Care Database (NPCD) 258
national pharmacovigilance systems 180–2
national postmarketing safety databases 182–3
National Research Council 7
National Research Ethics Service 311
National Science Foundation 837
National Survey on Drug Use and Health (NSDUH) 705–6, 714
National Transportation Safety Board 9
National Vital Statistics System (NVSS) Multiple Cause of Death data files 708, 714
Nationwide Emergency Department Sample (NEDS) 707
natural language processing (NLP) 117, 691
NAVIPPRO® *see* National Addictions Vigilance Intervention and Prevention Program (NAVIPPRO®)
NBIR see National Breast Implant Registry (NBIR) 503
NCCMERP *see* National Coordinating Council on Medication Error Reporting and Prevention (NCCMERP)
NCCN Task Force *see* National Comprehensive Cancer Centre (NCCN) Task Force
NCDR *see* National Cardiovascular Data Registry
NCI *see* National Cancer Institute
NDA *see* New Drug Applications
NDC *see* National Drug Codes
NDI *see* National Death Index
NDTI *see* National Disease and Therapeutic Index
needs assessment 128, 129
negative control object 1036
negative control precipitant 1035
negligence 74, 141, 142–3
NEISS *see* National Electronic Injury Surveillance System

NEISS-CADES *see* National Electronic Injury Surveillance System: Cooperative Adverse Drug Events Surveillance System (NEISS-CADES)
NEST *see* National Evaluation System for health Technology (NEST) Coordinating Center
nested case–control studies 230, 259, 328, 542–4, 1035, 1082–4, 1086
net benefit 847, 849
net clinical benefit (NCB) measures 876, 883–4
Neuro-QoL 1073
New Drug Applications (NDA) 101, 605
new drug approval processes 901
next-generation sequencing (NGS) 762, 766
NHAMCS *see* National Hospital Ambulatory Medical Care Survey
NHDS *see* National Hospital Discharge Survey
NHI *see* Normative Health Information
NHS *see* National Health Service
NHSBSA *see* National Health Service Business Services Authority (NHSBSA)
NHSN *see* National Healthcare Safety Network
NIAMS *see* National Institute of Arthritis and Musculoskeletal and Skin Diseases
NICE *see* National Institute for Health and Clinical Excellence
NIH *see* National Institutes of Health
nivolumab 566
NLP *see* natural language processing
NNT *see* number needed to treat
nomogram 60
nondiagnosis-linked databases 385
nondifferential misclassification 953–5
noninterventional studies (NIS) 777, 782
nonpill formulations 1000–1
nonresponse bias 329
nonsteroidal antiinflammatory drugs (NSAIDs)
  beneficial drug effects 818
  confounding data 1097
  data mining 680
  data sources 369
  drug utilization 376, 380, 392, 395
  encounter databases 230
  inpatient databases 296
  large simple trials 802
  legislative instruments 141, 144
  metaanalyses 905, 908–9, 928–9
  validity of data 953, 966, 967–9, 973, 974–9
nonuser comparison groups 1080
NSDUH *see* National Survey on Drug Use and Health
number needed to harm (NNH) 881–2
number needed to treat (NNT) 881–2, 1017
Nurses' Health Study 869
NVSS *see* National Vital Statistics System (NVSS)

O

object triggered concomitancy 1040
Observational Health Data Sciences and Informatics (OHDSI) 621, 632, 682, 933, 934
Observational Medical Outcomes Partnership (OMOP) 116, 618, 621, 624, 628, 683, 689, 933, 934
  OMOP CDM 628, 688
observer-reported outcomes 1063
obstetrics *see* pregnancy
odds ratios (OR)
  clinical epidemiology 54
  comparative effectiveness 301
  data mining 185, 685
  metaanalyses 912–13, 917–19, 926–7, 929, 930
  pharmacoepidemiologic research 759
  pharmacokinetic gene–drug interactions 751
  sample size 65
  validity of data 953
off-label usage 10, 148, 170, 185, 327, 334, 524, 527, 600, 611
OHDSI *see* Observational Health Data Sciences and Informatics (OHDSI)
olanzapine (Zyprexa®) 141
OMOP *see* Observational Medical Outcomes Partnership
oncology 756–7
one-group pre–post design 414
one-group, post-only design 414
Ontology of Adverse Events (OAE) 690, 691
Opana® ER (oxymorphone extended-release) 715–16
open-ended questions 961–2
opinion leaders 421–2
opioid abuse *see* drug abuse
opportunity cost 846
optimization of therapy 4
OPV *see* oral polio vaccine
oral contraceptives 916, 1002, 1018
oral polio vaccine (OPV) 437, 438, 440, 453,456, 457, 461
orphan drugs 98, 127
OrthoCRN 511
orthogonality 684,685
osteoporosis 822–3
OTC *see* over-the-counter
outcome
  delayed drug effect 368
  no medical attention 368
outcome data
  beneficial drug effects 820
  encounter databases 227
  event monitoring 310, 312, 316, 327–8
  medical devices 513
  medication errors 1052, 1055
  metaanalysis 916
  patient compliance/adherence 994
  vaccines 444–5, 466–7
  validity of 364–5
outcomes research *see* comparative effectiveness research
out-of-plan care 363
outpatient diagnoses
  data on 364
  data resources 365
  electronic databases 206, 207
  encounter databases 228
  medication errors 1054
  validity of data 966–9, 972, 973
overdose
  risk factors 1053
  risk management 586
overreporting 964, 966, 996, 999
over-the-counter (OTC) medications
  drug abuse 708
  medication errors 1049
  randomized controlled trials 802, 804, 809
  risk management 586
oxcarbazepine 322–4

p

pack size 607
packaging
  defects 506
  EDMs 999
  evaluation 595
  medication errors 585
  size of 1000
Paediatric Active Enhanced Disease Surveillance (PAEDS) system 449
PAEDS *see* Paediatric Active Enhanced Disease Surveillance (PAEDS)
pain 993
pandemic H1N1 (pH1N1) virus
  vaccines 437, 438, 445, 448, 449, 451, 454, 456, 457, 460, 463, 464, 466, 467
pantoprazole 1036
paracetamol 586, 601–2
parallel trials 917
Parent Project Muscular Dystrophy 886
paroxetine (Paxil®) 141, 143, 1030, 1039
PAS *see* postapproval studies
passive surveillance 447,448–9, 456, 458, 473, 506
  see also spontaneous reporting
patent rights 838
patents 140, 155–8
  see also intellectual property law
patient-centered care 1061
Patient-Centered Clinical Research Network (PCORnet) 92, 573, 620, 621, 624, 631, 807
patient-centered health research *see* comparative effectiveness research
Patient-Centered Outcomes Research Institute (PCORI) 92, 640, 644, 657, 1072, 1073
Patient-Centered Outcomes Research Trust Fund (PCORTF), 640
patient compliance/adherence 991–1005
  adherence metrics 994
  beneficial drug effects 814
  clinical problems 992–3
  clinical trials 1001
  comparative effectiveness studies 1001
  confounding data 1101
  current solutions 994–1004
  data analysis 994, 1001
  definitions 991–2
  drug concentrations 100
  drug utilization 380, 383, 387
  electronic drug monitors 995, 999
  future directions 1004–5
  measurement challenges and techniques 993–4
  medication diaries 998–9
  methodologic problems 993–4
  nonpill formulations 1000–1
  pharmacy dispensing data 995, 996–8
  pill counts 998
  prediction 1004
  primary 1000
  self-report data 995–6, 1000–1
  statistical analysis 1002–3
  temporal effects 1003–4
patient engagement 1061–74
  individualized assessment and treatment plans 1068–9
patient information leaflets (PIL) 592–3, 607, 1014–16
patient-oriented indicators 395–6
Patient Package Inserts 592, 605, 1018, 1020
Patient Protection and Affordable Care Act (PPACA) 639–40
patient registries *see* registries
patient-reported outcomes (PROs) 1061–74
  barriers to measuring 1069
  bookmarking and scale judgement of IRT-based measures 1071–2
  change in perspective 1072
  clinical problems 1067–9
  clinical trials 1063–4
  completion and results review 1067–8
  data collection 1066
  future of 1073–4
  in guiding interventions 1069
  methodologic problems 1069–72
  mobile health (mHealth) technology 1066
  new therapeutic strategies 1065–6
  routine case 1064–5
  selection and implementation into practice 1072–3
  selection, score interpretation and interventions 1068
  statistical methods 1071
  within-person change measurement 1070–2
Patient Reported Outcomes Measurement Information System (PROMISR) 1062
patient router files 227
patient variation 513
PBM *see* Pharmacy Benefit Management
PBRERs *see* periodic benefit–risk evaluation reports (PBRERs)
PCORI *see* Patient-Centered Outcomes Research Institute
PCORnet *see* Patient-Centered Clinical Research Network (PCORnet)
PDUFA *see* Prescription Drug User Fee Acts
Pedianet 253, 254, 255, 262, 264, 544
Pediatric Committee (PDCO) 530
Pediatric Health Information System (PHIS) 290, 292–4, 544
  applications 296, 299–300, 302
  strengths 294
  weaknesses 295
pediatric investigation plans (PIP) 98, 106, 111, 334, 530
Pediatric Regulation (2007) (EU) 530
Pediatric Research Equity Act (2003) (PREA) 530
pediatrics *see* children
PEDSnet 544
Peginesatide 192–3
PEM *see* prescription–event monitoring
People-Powered Research Networks 621
perceptions of risk 78–9
performance outcomes 1063
periodic benefit–risk evaluation reports (PBRERs) 867, 870
Periodic Safety Update Reports (PSUR) 447, 588
persistence 991–2
  definition 86
personal identification numbers (PIN) 229, 386
personnel 1117–18
Peto method 914, 917, 923, 929
Pfizer Pharmaceuticals 74, 348
PGRN *see* Pharmacogenetics Research Network

Pharmaceutical Benefits Advisory Committee 640
Pharmaceutical Manufacturers Association 775–6
Pharmaceutical Research and Manufacturers of America Benefit-Risk Action Team (PhRMA BRAT) framework 870–1,872
pharmacodynamics 37–40
  children 527
  definition 4
  elderly populations 33
  molecular pharmacoepidemiology 752–3
pharmacoeconomics 837–60
  behavioral economics 841–3
  clinical problems 837–9
  currently available solutions 845–6
  economic evaluation and drug development process 845–6
  economic valuation 846–7
  future of 857–60
  health economics 839–45
  Imatinib case study 852–7
  methodologic issues 846–57
  moral hazard 839–41
  perspective of analysis 851
  pharmaceutical products 837–8
  types of analysis 847–50
  types of costs 850–1
pharmacoenvironmentology. 786
Pharmacoepidemiologic Research on Outcomes of Therapeutics by a European Consortium *see* PROTECT
pharmacoepidemiologic studies
  acceptability of risks 77–9
  benefits of 71–6
  costs of 71
  reasons to perform 71–6
  risk tolerance 76–9
  risks of conducting 71
  safety versus risk 76
pharmacoepidemiology
  definition 3–4
  vs epidemiology 5
  historical background 6–13
  potential contributions of 15–18
pharmacogenetics 746, 748, 749–50
Pharmacogenetics Research Network (PGRN) 1117
pharmacogenomics 749–50
  clinical pharmacology 40
  data mining 676
  pharmacogenetics and 749–50
  vaccines 442, 452
pharmacokinetics 29–37
  absorption 29
  children 527
  definition 4
  drug interactions 37
  drug utilization 380
  elderly populations 32–4
  elimination 31–2
  metabolism 30–1
  molecular pharmacoepidemiology 749–50
  organ impairment 36
  pediatrics 34–5
  pregnancy 35–6, 526
  risk management 588
  special populations 32–7
  volume of distribution 29
pharmacology, definition 4
pharmacovigilance 777–8
  causation 730
  definition 5
  disaster-driven 678–9
  national pharmacovigilance systems 180–2
  postmarketing safety databases 182–5, 187–9, 196
  risk management 606–7, 611
  signal detection 183–5, 187–8, 677, 678
  see also spontaneous reporting
pharmacovigilance plans 582
pharmacovigilance reporting systems 169–96
  AE/ADR reporting 171–2, 188
  AE/ADRs, definitions 170–1
  applications 190–6
  developing countries 195
  drug safety evaluations 107, 109, 117–18
  future 196
  limitations 189–90
  medication errors 1048
  national pharmacovigilance systems 180–2
  nonuniformity with time 189–90
  overview 174–5
  patient reports and healthcare professional reports 175–6
  postmarketing safety databases 182–3, 185, 187
  report characteristics 177–9
  report quality 189
  reporting ratios 186–7, 189
  reviews of case reports 185–6
  scope 188–9
  signal detection 183–5, 187–8
  social media 179–80
  strengths187–9
  underreporting 189
  vaccines 447–8, 449, 455, 458, 460
  see also case reports
Pharmacovigilance Risk Assessment Committee (PRAC) 344
pharmacovigilance system master file (PSMF) 100
Pharmacy Benefit Management (PBM) databases 258
pharmacy dispensing data 995, 996–8
PharmaNet 226
Pharmetrics® 360
PharmGKB 756
phase I metabolic reactions 30–1
phase II metabolic reactions 31
Phelan-McDermid Syndrome Data Network 621
phenylbutazone 7, 66
phenytoin 30, 31
PHI *see* protected health information
PHIS *see* Pediatric Health Information System
phocomelia 307, 773, 774
physician prescribing *see* prescribing practices
Physicians' Health Study 800
PICORI 644
pictographs 1012, 1014, 1016
pill counts 998
pill fatigue 993
PIM *see* potentially inappropriate medication
PIN *see* personal identification numbers
Pink Sheet 345
pioglitazone 75
PIP *see* pediatric investigation plans
piroxicam (Feldene®) 74, 75
placebo-controlled trials 656, 901, 921–2, 924, 931
PMC *see* postmarketing commitments
PMDPs 715

PML *see* progressive multifocal leukoencephalopathy
PMR *see* postmarketing requirements
POCCT *see* point-of-care clinical trial (POCCT)
point estimate of association 761, 961
point-of-care clinical trial (POCCT) 807
poison control centers (PCCs) 707
Poisson regression modeling 332–3
polychlorinated biphenyls (PCBs) 148
population admixture 760–1, 765
population at risk, concept 91
population based database 212, 363
population-based data
  causality 728
  drug utilization 390
  medical devices 503–4
  risk management 601
  see also registries
positive control precipitant 1039
positive predictive value (PPV) 216, 966, 974, 979
postapproval monitoring 85
postapproval studies (PAS)
  biopharmaceutical industry 106–8
  medical devices 511
  regulatory agencies 130–3
  risk assessments 589–90
  vaccines 461
postauthorization efficacy studies (PAES) 100
postauthorization safety study (PASS) (EU) 100, 337, 783–4
Post-licensure Rapid Immunization Safety Monitoring (PRISM) project 463–4
postmarketing commitments (PMC) 99
Postmarketing Drug Surveillance Research Unit 308
postmarketing event monitoring 308
postmarketing requirements (PMR) 99
postmarketing safety studies 308
postmarketing study, costs of 71
postmarketing surveillance (PMS) 100, 307
  beneficial drug effects 816, 826
  clinical epidemiology 51, 53
  costs 71
  health-care systems 100
  hypothesis generating studies 75
  regulatory factors 72–3
  sample size 68, 69
  study designs 308
  see also pharmacovigilance; spontaneous reporting
potentially inappropriate medication (PIM) 299
power
  in large simple trials 798
  in randomized controlled trials 797
PPACA *see* Patient Protection and Affordable Care Act
PPI *see* proton pump inhibitors
PPV *see* positive predictive value
practolol 7, 308, 775
pragmatic clinical trials (PCTs) 793
pravastatin 1039
prazosin 16, 51, 66, 72, 75, 813
PREA *see* Pediatric Research Equity Act
preapproval review process 170
preapproval safety studies
  biopharmaceutical industry 103–6
  risk assessments 589
  spontaneous reporting 170
  see also clinical trials
PRECEDE model 419
precipitant triggered concomitancy 1040
Precision Medicine Initiative 843
preference studies 878, 884,885
Preferred Reporting Items for Systematic reviews and Metaanalyses (PRISMA) 648, 902, 906–7, 910
pregnancy 524–47
  automated healthcare databases 543
  bias in samples 525
  biology and epidemiology 525–6
  case–control studies 545
  clinical practice 530–1
  confounding 537–8
  currently available solutions 541–6
  drug safety evaluations 108–11
  drug utilization 526–7
  ethics and regulations 528–9
  exposure ascertainment, timing, and misclassification 532–3, 533–4
  future 546–7
  medication adherence 526–7
  medication errors 1049
  methodological problems 531–41
  newer designs 545
  outcome and ascertainment 534–6
  pharmacokinetics 36
  population definition 531
  prospective cohort 541
  registries 541–2
  risk management 595
  risk tolerance 78
  sample size requirements and challenges 532
  selection bias 538–40
  self-controlled designs 545
  teratogenicity 528, 529
  treatment responses and patterns 526–7
  see also birth defects
premarketing studies 17–18
premarketing trials, shortcomings of 84–5
Premier Healthcare Database (PHD) 290, 294, 299, 301, 302
pre–post with comparison group design 416
prescribed daily dose (PDD) 389
prescribing practices 411–29
  academic detailing 420–1
  audit and feedback systems 422–3
  clinical problems 412–13
  conceptual framework 419
  detecting outcome effects 418–19
  drug utilization 380, 381–2, 383. 385–9
  educational materials and guidelines 419–20
  effectiveness of interventions 419
  ethical and legal problems 417–18
  evaluating and improving 411–29
  financial incentives and formulary interventions 426–7
  future 427–9
  group detailing 421
  group education 421
  internal validity 414–16
  logistical issues 417
  medication errors 1047, 1049
  methodologic problems 413–19
  multimedia campaigns 425
  opinion leaders 421–2
  regression toward the mean 416

prescribing practices (*cont'd*)
reminders and computerized decision support 424, 428
risk management 593
small group learning 421
study design 414
unit of analysis 416–17
validity of data 951, 972
Prescription Card Services (PCS) 776
prescription drug monitoring programs (PDMPs) 704, 711, 712
Prescription Drug User Fee Acts (PDUFA) 9–10, 11, 16, 99, 1063
Prescription Event Monitoring (PEM) 11, 308, 363, 365, 367, 368, 369
see also event monitoring; modified PEM; specialist cohort event monitoring (SCEM)
prescription medications 586, 591, 593, 599
Prescription Pricing Division (PPD) 309
preterm births 34, 526, 533, 538, 540, 543
pretesting and validation 597
prevalence studies, sample size 65–6, 68, 69
primary analysis 801
primary data collection 342–51
applications 349–50
biobanks/specimen banks 347
clinical assessments 343–4
clinician- or site-reported outcomes 346
description 346–7
endpoint assessment/adjudication 343
future 350–1
hybrid or enriched designs 345
limitations 348–9
medical devices 345
patient-reported measures 368
patient-reported outcomes 344, 346
quality of data collection 347–8
randomization 342–3
rare populations 344
registries 346–7
research questions requiring 342–5
special requirements, controlled distribution products 345
strengths 347–8
vaccine safety 344–5
privacy and confidentiality 775, 778–9, 780, 781–2
private insurance databases 220–2
probabilistic linkage 217, 508, 573, 653
probabilistic methods 735–8
PROBE 806
procedure claims 230, 570
product liability 730–1
in Europe 150–2
lawsuits 140–8
legal theory of 141–8
effect of regulation on, in US 148–50
product lifecycles
regulatory agencies 126–7
risk management 582, 587, 588
spontaneous reporting 170
product quality 178, 584, 585
product use information 178
productivity costs 845, 847, 850, 851
progressive multifocal leukoencephalopathy (PML) 114–15, 190–1, 884
PROLIA® 1020
PROMIS 573
PROMISR (Patient-Reported Outcomes Measurement Information System) 1073
propensity score (PS) 537, 629, 820, 1086–7, 1091–4, 1095
propensity score calibration (PSC) 1099–100
propensity score matching 711
proportion of days covered (PDC) 997
proportional reporting ratio (PRR) 334, 685
propoxyphene 8, 299, 421
prospective studies
data collection 343–4
electronic databases 205
PROTECT study 343, 348, 349, 350, 385, 620, 630
protected health information (PHI) 781
protocol development 443–4
proton pump inhibitors (PPI) 116, 230, 412, 1030, 1036, 1039, 1040
PROVALID study 344
proxy adjustment 1095–6
PRR *see* proportional reporting ratios
PS *see* propensity score
PSC *see* propensity score calibration
PSMF *see* pharmacovigilance system master file (PSMF)
PSUR *see* Periodic Safety Update Reports
public reporting 188
publication bias 903–5, 918–20, 930, 931
publicity 190
Pure Food and Drug Act (1906) 6
P-values 54

q

QALY *see* quality-adjusted life-years
quadrivalent live attenuated influenza vaccine (QLAIV) 337
qualification stickers 602–3
qualitative drug utilization studies 381
quality-adjusted life-year (QALY) 844, 848, 849
Quality and Outcomes Framework (QOF) 256
quality assurance 212, 231, 256, 259, 294, 387, 414, 585
quality control 223, 585, 685, 757, 783
quantitative drug utilization studies 381
Quantitative Sciences in the Pharmaceutical Industry (QSPI) Benefit–Risk Working Group 886
quantitative strength of associations 48
quasi-experimental study design 418
questionnaire design
case–control studies 52
cohort studies 53
drugs of abuse 706
electronic data capture 337
electronic health record databases 256
event monitoring 311–14, 314, 318, 319, 325, 327–9
PEM 335
risk management 597
validity of data 961–2, 964–6

r

race/ethnicity
molecular pharmacoepidemiology 760–1, 764
validity of data 959, 962

RAIDAR *see* rare and iatrogenic adverse reactions
random exposure variation 1096–9
random-effects models 913–14, 916, 924
randomized controlled trials (RCT) 792–809
  beneficial drug effects 818, 822, 828
  benefit–risk assessment 869
  biopharmaceutical industry 104–5
  children in 782
  clinical epidemiology 44, 50, 52, 55
  clinical problems 794–5
  comparative effectiveness research 639, 655
  concealed (masking) vs known treatment 795–7
  confounding by indication 794, 798, 807, 809
  data analysis 801–2
  data collection 797, 801
  data elements 797
  data mining 677, 678
  future 807–9
  medical devices 501–2
  metaanalyses 897–9, 908–9, 924, 928, 930–1
  methodologic problems 795–7
  postapproval 793
  power 797
  prescribing practices 413, 414, 419, 421, 423, 427
  sample size 797
  simple protocols 795
  see also large simple trial (LST)
Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) study 872
rapid-cycle analyses (RCA) 685
  vaccines 455–6, 462–3
rare and iatrogenic adverse reactions (RAIDAR) 317
rare diseases52, 54, 66, 530, 541, 544, 589, 629, 777, 886
rare events 916–17
RCA *see* rapid-cycle analyses
RCTs *see* randomized controlled trials
READ codes 254, 626
Reagan–Udall Foundation 933–4
recall bias
  validity of data 960–8, 969
rechallenge 726, 731, 732, 734
RECORD statement 684
Recovery and Reinvestment Act 502, 639
registries
  data collection 346–7
  drug safety evaluations 101, 103, 105–6
  medical devices 509–11
  pregnancy 541–2
Registry in Glaucoma Outcomes Research (RiGOR) study 349
regression analyses 631, 709, 878
regression toward the mean 416, 819
regulatory agencies 126–36
  advancing pharmacoepidemiology 134–6
  assessing impacts of regulatory actions 132–4
  collaboration 134–6
  future developments 1112, 1119–20
  integrating data from multiple sources 127
  monitoring postapproval safety 131–2
  needs assessment 126–7
  orphan drugs 127
  planning drug development programs 127–30
  postapproval safety studies 130–1
  preapproval clinical safety data 130
  product lifecycles 126–7
  use pattern assessments 132
regulatory benefits 72–3
regulatory factors
  academia 111–12, 1112–13
  assessing impacts of regulatory actions 132–4
  biopharmaceutical industry 98
  causation 731–2
  ethics 773, 777–8, 780, 783–4
  medical devices 497–501, 515–16
  metaanalyses 930–1
  pharmacoepidemiological studies 72–3
  risk management 604–12
reimbursement policies
  drug utilization research 379
  reinforcement strategies 419
relative risk
  adverse outcome features 76–7
  clinical epidemiology 48, 53–4
  confounding data 1081, 1100
  metaanalyses 912, 913, 915, 917, 918, 920, 925, 927–9, 932
  risk evaluation studies 1012
  sample size 61–4
  validity of data 953–4
relative size of databases 360, 363
relative speed of databases 360
reliability
  measurement error 949
  quantitative measurement 951–2
reminders 424, 428
REMOTE 348
REMS *see* risk evaluation and mitigation strategies
renal function 31–2, 34
replicate association studies 763
report quality 189
reporting odds ratio (ROR) 685–6
reporting ratios 186–7, 189
research collaborations *see* collaboration
Research Data Assistance Center (ResDAC) 223
research questions
  beneficial drug effects 819
  confounding by indication 819
  large simple trials 798
research subjects 772, 780, 782
reserpine (Harmonyl®) 147
resource-poor countries 396
resource use 853, 854, 855
response order 961, 965
response shift 1072
Restatements of Torts 141–2
restriction criteria 1094–5
Retropro trial 806
retrospective studies
  automated healthcare databases 542–4
  clinical epidemiology 56
  confounding data 1100
reviews of case reports 185–6
Reye's syndrome 425, 802,803
rhabdomyolysis 1041
rheumatoid arthritis (RA)
  confounding data 1084
  patient-reported outcomes 1070
Rheumatology Informatics System for Effectiveness (RISE) 573
rights of patients/subjects 772, 777, 787

RiGOR *see* Registry in Glaucoma Outcomes Research (RiGOR) study
risk assessment 99, 588–90
risk–benefit profiles *see* benefit–risk profiles
risk communication 585, 591–6, 1012–13
risk difference (RD)
  confounding data 1098
  Medication Guides 1020
  metaanalyses 912, 918, 927, 929, 932
risk evaluation and mitigation strategies (REMS)
  biopharmaceutical industry 99, 112–13
  opioid analgesics 701
  risk evaluation studies 1019–21
  risk management 593, 596, 600, 604–5
risk evaluation studies 1010–24
  clinical problems 1011–12
  communication of risks 1012–13, 1021–2
  current solutions 1014–22
  direct-to-consumer advertising 1022
  future developments 1022–3
  fuzzy-trace theory 1023
  health professional/consumer medication information 1014–21
  Internet 1022
  methodologic problems 1012–13
  point estimate of association 1023
  risk evaluation and mitigation strategies 1020–1
  social media 1022
risk-interval analysis 463–4, 468–70
risk management 12, 581–14
  activities 587–99
  applicability for all medicines 586–7
  biopharmaceutical industry 99–100, 112–15
  clinical problems 582–7
  current solutions 604–12
  effectiveness 587, 589
  future developments 612–14
  iterative and multidisciplinary nature 587–8, 598–9
  medication use system complexity 582–3
  methodologic problems 599–604
  product lifecycles 585–6
  regulatory factors 604–12
  risk assessment 588–90
  risk communication 591–6
  risk mitigation 590–9, 601–2
  scientifically driven 585
  sources of risk 583–5
risk management plans (RMP) 98, 99–100, 111, 309, 607, 1014, 1015
risk measurement 76
Risk Minimization Action Plans (RiskMAP) 99
risk minimization *see* risk mitigation
risk mitigation 590–9, 610–14
  biopharmaceutical industry 99, 105, 106, 112–15
  evaluation 596–9, 601–2
risk set sampling 1035
risk tolerance 76–9
  adverse outcome 76–7
  exposure 77–8
  perceptions of evaluator 78–9
RiskMAP *see* Risk Minimization Action Plans
rituximab 564
Rivaroxaban Observational Safety Evaluation (ROSE)SCEM study 329
RMP *see* risk management plans
rofecoxib 75, 141, 144, 146
rosiglitazone 75, 144, 923
route-of-abuse profile 710
routine risk minimization 608
RUCAM scoring method 735

S

SAE *see* serious adverse events
safety assessment 75
Safety Evaluation of Adverse Reactions in Diabetes (SAFEGUARD) 621
Safety of Nonsteroidal Anti-inflammatory Drugs (SOS) 621
Safety Planning, Evaluation and Reporting Team (SPERT) 900, 932
safety profiles 75, 900, 901
safety registries 345
Safety Report Cards 345
safety signal detection 677
sample size 60–70
  beneficial drug effects 822
  case–control studies 60, 62, 65–6, 68–9
  case series 66–8
  cohort studies 61–4, 68–9
  drug safety evaluations 104
  electronic databases 70
  event monitoring 316–18
  large simple trial 797
  medical devices 505–6
  randomized controlled trials 797
  vaccines 441
sampling errors 916, 955
Saskatchewan Health database 974, 1083
  validity of data 976, 977
satisficing 958, 961, 964, 965
scale judgment method 1072
SCCS *see* self-controlled case series
schizophrenia 885
scientific inference 45
scientific method 44–5
scored algorithms 735
screening 75
secondary database studies 1080, 1099
secondary nonadherence 87
secretion 32
SEER *see* Surveillance, Epidemiology, and End Results
selection bias 52
  children 540–1
  confounding data 1090
  electronic health record databases 259
  encounter databases 212
  event monitoring 329
  pregnancy 110, 538–40
  vaccine safety 468, 469, 470
  validity of data 955
  voluntary enrollment 109
selective serotonin reuptake inhibitors (SSRI) 141, 143, 148
self-administered questionnaires 958, 965, 966
self-controlled case series (SCCS) 468–9, 1037, 1038
self-management support 1061, 1065
self-report data
  patient compliance/adherence 995–6
  validity 949, 955–69
  see also spontaneous reporting

SEM *see* standard error of measurement
sensitivity analysis
  confounding data 1100–1
  pharmacoeconomics 850
  validity of data 950, 953, 954, 955, 966, 971, 972
Sentinel Initiative 116, 136, 620, 680, 776
  applications 622–3
  automated healthcare databases in pregnancy 543
  common data models 624, 627–8
  data mining 680–1, 683, 688, 692
  future developments 630–2, 1116
  metaanalyses 933
  modular analyses 369
  replication in 683
  research questions 365
  risk evaluation studies 1019
  risk management 600
  venous thromboembolism (VTE) 681
  see also Mini-Sentinel
sepsis 301–2
serious adverse events (SAE) 571
  large simple trial 803
  randomized controlled trials 802, 803
  risk management 584, 587, 595, 600
  spontaneous reporting 183
serotonin syndrome 1032
Shanghai Center for Adverse Drug Reaction Monitoring 779
Shewhart charts 507
Sibling discordance studies 545
side effects *see* adverse events/adverse drug reactions
SIDIAP 253–4, 255, 262, 263, 264
Sierra Leone, mass drug administration in 195–6
sigmoid concentration effects 38
signal detection
  metaanalyses 928–9
  spontaneous reporting 183–5, 187–8
  vaccines 443, 454–6
  see also Sentinel Initiative
signal generation 176, 318, 325, 330, 333, 455
signal refinement 676, 678, 680
sildenafil 8, 102, 333
similarity of data 922
simple stratification 331–2
simultaneous comparisons 921–2, 930
single nucleotide polymorphisms (SNP) 40, 748, 749, 755, 761–2
single-group cohort design 329–30
SIP *see* Sickness Impact Profile
Sistema de Información para el Desarrollo de la Investigación en Atención International Classification of Primary Care (ICPC) 254
SJS/TEN *see* Stevens–Johnson syndrome/toxic epidermal necrolysis
Slone Epidemiology Center 11, 545
Slone Epidemiology Unit 360
small group learning 421
SNOMED *see* Systemized Nomenclature of Medicine
SNP *see* single nucleotide polymorphisms
social desirability 996
social media 179–80, 688–90, 1022
Social Security Administration (SSA) 258, 851
Society for Medical Decision Making (SMDM) 886
sofosbuvir 844
solicited report 173–5
SOS 266
South Korean: Health Insurance Review and Assessment Data 229
sparse additive models 505
sparse prior distributions for regression coefficients 505
special populations
  drug safety evaluations 108–12
  medication errors 1051
  pharmacokinetics 32–7
  see also children; elderly populations; pregnancy
specialist cohort event monitoring (SCEM) 308
specific causation 146
specificity 48, 950
SPERT *see* Safety Planning, Evaluation and Reporting Team
spontaneous reporting methods 308
spontaneous reporting systems (SRSs) 363, 367, 368–9, 677
SSA *see* Social Security Administration
SSRI *see* selective serotonin reuptake inhibitors
standard error of measurement 1071
standard of care 343
standardized clinical assessment protocols/centers 451–2
standardized mean differences (SMD) 924
standardized mortality ratio (SMR) 333
Stanford Health Assessment Questionnaire (HAQ) 968
statin-associated rhabdomyolysis 175
statins 1041, 1093–4
statistical inference 45
statistical power
  metaanalyses 922–3
Stevens–Johnson syndrome 186, 188, 209, 333, 334, 586, 677, 736, 737, 740, 746, 808, 959
stock outs 993
stopping medication 325, 414
STREGA statement (Strengthening the Reporting of Genetic Association studies) 764
strength of associations 48–9
STROBE statement (Strengthening of Observational Studies in Epidemiology) 764–5
stroke 8, 88, 129, 191–2, 610, 872, 878
study design
  clinical epidemiology 48–50, 50–5
  confounding data 1079–101
  drug safety evaluations 101–2
  prescribing practices 414
  vaccines 445–6, 467–70
study power 797, 798
study subjects 44–5
subacute myelooptic-neuropathy (SMON) 7
subgroup analysis 801–2, 924–6
substance use disorder (SUD)
  insurance claims 707–8
suicidality 924–5, 931
sulfanilimide 6
sulfonylurea antidiabetic agents 1030
Summary of Product Characteristics (SmPC) 319, 591, 592–3, 594
supplementary information 17–18
suprofen 7
surgeon and center variation 513
surrogate markers 128–30

surveillance
active 782
case–control 360, 364, 365 367, 369, 370
ethics 774, 775
postapproval 461
Surveillance, Epidemiology, and End Results (SEER) program 346–7
Swedish Prescribed Drug Register 386
synergy index 762
Syntex 74
System to Manage Accutane-Related Teratogenicity (SMART) 602–3
systematic reviews
medical devices 505
metaanalyses 897, 898, 902, 904, 906
spontaneous reporting 186
Systemized Nomenclature of Medicine (SNOMED) 392
systems approach 12, 613

t

T-2 translational research 411
Tagamet® 776
Taiwan
National Adverse Drug Reactions Reporting System 779
National Health Insurance program 779
National Health Insurance Research Database 228–9
tamoxifen 1030
Task Force on Research Specific to Pregnant Women and Lactating Women 529
Task-force in Europe for Drug Development for the Young (TEDDY) Network 266, 544
temporal trends
inpatient databases 296
patient compliance/adherence 1003–4
reporting 189–90
spontaneous reporting 189–90
validity of data 965, 966
venous thromboembolism (VTE) 296
teratogenesis *see* birth defects
terminal illness 78
thalidomide 6, 307, 534, 773, 774
The Health Improvement Network® (THIN®) 207, 234, 252–7, 262, 263, 328, 363, 544
therapeutic risk management 1113, 1119
thiazolidinediones 75
THIN® *see* The Health Improvement Network®
thiopurine methyltransferase (TPMT) 751
thiopurines 565
third-party licenses 156
time period covered by a database 215
time savings 926–8
time sequence of associations 48
time-series regression analysis 709
time-to-event analysis 325
time to onset 325
TNFi therapy 567
tobacco-related diseases 78
tocilizumab 565, 566, 569
tolmetin 66, 73, 75
topical treatments 1000–1
topiramate 195
tort law 140–52
TOUCH risk minimization system 114
toxic epidermal necrolysis 186, 746
toxic shock syndrome 76–7, 78
toxicology 74–6
Toxicology Investigators Consortium (ToxIC) 705
TPMT *see* thiopurine methyltransferase
Transcatheter Valve Therapy (TVT) registry 511
Transformed-MSIS (T-MSIS) 223
transparency 786
trastuzumab 753, 756, 757
treatment efficacy vs. effectiveness 642
Treatment Episode Data Set (TEDS) 706
TreeScan 464
triazolam 7
trim-and-fill method 920
troglitazone (Rezulin®)
law and 141, 144
liver disease and 8, 75
Truven Health MarketScan Database 290, 294, 295
tuberculosis 564, 993
Tufts Cost-Effectiveness Analysis Registry 882
21st Century Cures Act (2016) 10, 85, 533
Type A reaction 4, 77
Type B reactions 4–5, 77
Type I/II errors
data-mining 473, 682
drug–drug interactions 1032, 1043
event monitoring 318
metaanalyses 926–7
molecular pharmacoepidemiology 752, 757, 759–60, 763–5
sample size 61, 63, 65, 330
sequential testing 473
signal detection 454
vaccines 454, 473

u

UDI *see* unique device identifiers
UGT *see* glucuronosyl transferase
UMC *see* Uppsala Monitoring Centre
unanticipated beneficial effects 814
unanticipated harmful effects 813
uncertainty
confounding data 1100
large simple trials 798
risk evaluation studies 1010
validity of data 955
underdosage 1053
underprescribing 85
underreporting
patient compliance/adherence 999
spontaneous reporting 189
validity of data 964–5
unexposed control groups 61–5, 68, 1099
Unique Device Identification (UDI) 498
unique device identifiers (UDI) 504
United Kingdom electronic health record databases 252–7
United States government claims databases *see* government claims databases
UnitedHealth Group 221
units of analysis 416–17
unpreventable risks 584
unpublished studies 918–20
unstructured clinical judgments 728, 733–4

Up-to-Standard (UTS) date 256
Uppsala Monitoring Center (UMC) 183, 449
Urabe strain 440, 444, 453
use pattern assessments 132
ustekinumab 569

V

VA *see* Department of Veterans' Affairs
Vaccine Adverse Event Reporting System (VAERS) 448–9, 454–60, 681
Vaccine Adverse Event, Surveillance and Communication (VAESCO) 454
Vaccine Injury Compensation Program (VICP) 465
Vaccine Injury Table 440
Vaccine Research Centers (VHC) 442
Vaccine Safety Datalink (VSD) 92, 444, 455–6, 462–3, 467, 469, 618, 620–1, 681
vaccine-preventable diseases (VPD) 437, 441, 453
vaccines 437–73
  ad hoc epidemiologic studies 461
  analytic methods 467–70
  automated large, linked databases 461–5
  beneficial drug effects 827
  causation 444, 452–4, 729
  classifications and case definitions 450–1
  clinical issues 441–2
  clinical trials 460–1
  currently available solutions 446–71
  data analysis 445–6
  definitions and protocols 443–4
  exposure data 444, 465–6
  exposure misclassification 444
  future 471–3
  injury compensation 465
  large immunization campaigns 456
  lessons learned 456–60
  methodologic problems 443–6
  methodologies 465–71
  observational epidemiologic outcome data 444–5, 466–7
  policy issues 438–41
  postapproval surveillance studies 461
  postlicensure trials 447, 460–1
  prelicensure trials 446–7, 460
  signal detection 443, 454–6
  standardized clinical assessment protocols/centers 451–2
  study design 445–6, 467–70
VAD *see* ventricular-assist devices
VAERS *see* Vaccine Adverse Event Reporting System
VAESCO *see* Vaccine Adverse Event, Surveillance and Communication
validity of data 948–80
  accuracy 959–62, 969, 965–6, 970, 972–4
  administrative databases 949–51, 969–79
  clinical problems 948–9
  comparator selection, 955, 959–60, 970
  current solutions 955–79
  demographic data 962, 965, 966, 969
  diagnosis data 963–79
  future developments 979–80
  measurement error 949, 952–5
  medical records 959–61, 965–6, 969,970, 973–4, 979
  metaanalyses 921,933
  methodologic problems 949–79
  misclassification errors 952–3
  outcome data 364
  patient compliance/adherence 998
  patient population 962, 965–6, 972
  prescribing practices 414–16
  quantitative measurement 949–52
  questionnaire design 958–9, 961–2, 966–9
  recall bias 960–8, 969
  reliability 951–2
  self-report data 949, 955–69
valproic acid/valproate 37, 527
Value Added Medical Products (VAMP) Research Databank® 252
vedolizumab 564
venous thromboembolism (VTE)
  anticoagulants for 823–4
  case–control studies 52
  cohort studies 53
  death rates from 51–2
  HPV vaccination and 463, 681
  oral contraceptives association 56, 116, 322
  population-based database and 363
  randomized clinical trials 55
  temporal trends 296
  time to onset 325
ventricular tachycardia 1089
ventricular assist devices (VAD) 513
verbal judgments 734
verification 949
Veterans Health Information Systems Technology and Architecture (VistA) system 258
Veterans' Affairs database *see* Department of Veterans' Affairs
veterinary medicine 777
VHC *see* Vaccine Research Centers
vigabatrin 318, 331, 599
VigiBase 176, 180, 183, 194, 195
VIGOR *see* Vioxx Gastrointestinal Outcome studies (VIGOR) trial
Vioxx Gastrointestinal Outcome studies (VIGOR) trial 680
VIRGO study 348
Virtual Data Warehouse 618–20
Virtual Research Data Center (VRDC) 223
VISION initiative 511
vitamin K epoxide reductase complex 1 (VKORC-1) gene 755
volume data 382
volume of distribution 29
VPD *see* vaccine-preventable diseases
VRDC *see* Virtual Research Data Center (VRDC)
VSD *see* Vaccine Safety Datalink
VTE *see* venous thromboembolism

W

warfarin 754–5, 764, 1030, 1037, 1038, 1039, 1040
warning failures 142, 143–6
washout periods 1100
Waxman–Hatch Act 825
Web Ontology Language (OWL) 690
WEB-RADR 180
WHO Adverse Reaction Terminology (WHO-ART) 459, 690, 691
winner's curse phenomenon 763–4
within-subject studies 1087–8

Women's Health Initiative (WHI) 12, 55, 823, 869
Women's Health Study 799
World Health Organization (WHO) 6
  drug approval processes 14–15
  drug utilization 375, 376, 377, 388, 390–2, 396
  early legislation 6
  spontaneous reporting 170, 176, 183, 184, 186, 194, 195
  vaccines 440, 448, 449, 450, 454
Wyeth v. Levine 150, 159

y

Yasmin® 151, 322
Yellow Card system 176, 307, 308, 329

z

Z-statistics 919
  zidovudine 73, 78
Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) Study 655, 799, 804–5
ZODIAC *see* Ziprasidone Observational Study of Cardiac Outcomes